

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2001, 07:52:13 ; Search time 2960.24 Seconds  
(without alignments)  
4561.571 Million cell updates/sec

Title: US-09-105-117I-1\_COPY\_1421\_2293  
Perfect score: 873  
Sequence: 1 atgaacccattcaactgga.....tcgaggattgcggccttag 873

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_ba3:\*
- 4: gb\_in1:\*
- 5: gb\_in2:\*
- 6: gb\_in3:\*
- 7: gb\_cm:\*
- 8: gb\_ov:\*
- 9: gb\_pat1:\*
- 10: gb\_pat2:\*
- 11: gb\_ph:\*
- 12: gb\_pl1:\*
- 13: gb\_pl2:\*
- 14: gb\_pl3:\*
- 15: gb\_pl4:\*
- 16: em\_ba1:\*
- 17: em\_ba2:\*
- 18: em\_fun:\*
- 19: em\_htgo\_hum:\*
- 20: em\_htgo\_inv:\*
- 21: em\_htgo\_rod:\*
- 22: em\_htg\_hum1:\*
- 23: em\_htg\_hum2:\*
- 24: em\_htg\_hum3:\*
- 25: em\_htg\_hum4:\*
- 26: em\_htg\_hum5:\*
- 27: em\_htg\_hum6:\*
- 28: em\_htg\_hum7:\*
- 29: em\_htg\_hum8:\*
- 30: em\_htg\_inv1:\*
- 31: em\_htg\_inv2:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_rod:\*
- 34: em\_hum1:\*
- 35: em\_hum2:\*
- 36: em\_hum3:\*
- 37: em\_hum4:\*
- 38: em\_hum5:\*
- 39: em\_hum6:\*
- 40: em\_hum7:\*
- 41: em\_in:\*
- 42: em\_cm:\*
- 43: em\_or:\*

- 44: em\_ov:\*
- 45: em\_pat:\*
- 46: em\_ph:\*
- 47: em\_pl:\*
- 48: em\_ro:\*
- 49: em\_sts:\*
- 50: em\_sy:\*
- 51: em\_un:\*
- 52: em\_v1:\*
- 53: gb\_sts1:\*
- 54: gb\_sts2:\*
- 55: gb\_sts3:\*
- 56: gb\_sy:\*
- 57: gb\_un:\*
- 58: gb\_v1:\*
- 59: gb\_v12:\*
- 60: gb\_htg1:\*
- 61: gb\_htg2:\*
- 62: gb\_htg3:\*
- 63: gb\_htg4:\*
- 64: gb\_htg5:\*
- 65: gb\_htg6:\*
- 66: gb\_htg7:\*
- 67: gb\_htg8:\*
- 68: gb\_htg9:\*
- 69: gb\_htg10:\*
- 70: gb\_htg11:\*
- 71: gb\_htg12:\*
- 72: gb\_htg13:\*
- 73: gb\_htg14:\*
- 74: gb\_htg15:\*
- 75: gb\_htg16:\*
- 76: gb\_htg17:\*
- 77: gb\_htg18:\*
- 78: gb\_htg19:\*
- 79: gb\_htg20:\*
- 80: gb\_htg21:\*
- 81: gb\_htg22:\*
- 82: gb\_htg23:\*
- 83: gb\_htg24:\*
- 84: gb\_htg25:\*
- 85: gb\_pr1:\*
- 86: gb\_pr2:\*
- 87: gb\_pr3:\*
- 88: gb\_pr4:\*
- 89: gb\_pr5:\*
- 90: gb\_pr6:\*
- 91: gb\_pr7:\*
- 92: gb\_pr8:\*
- 93: gb\_pr9:\*
- 94: gb\_rol:\*
- 95: gb\_rol2:\*
- 96: gb\_in4:\*
- 97: gb\_pr10:\*
- 98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
1	873	100.0	993	9	AX063767	AX063767 Sequence	
c	2	873	100.0	2374	2	CGLYSEG	X96471 C. glutamicu
c	3	873	100.0	2374	9	A93933	A93933 Sequence 2
c	4	140.6	16.1	33285	3	SC5F8	AL357613 Streptomy
5	122.6	14.0	14844	3	MBU34849	U34849 Mycobacteri	
6	122.6	14.0	15239	10	I86263	I86263 Sequence 17	
c	7	121	13.9	38500	3	MTCY39	Z74025 Mycobacteri
c	8	108.6	12.4	12070	1	AB011413	AB011413 Streptomy



CGLYSEG/c  
 LOCUS 2374 bp DNA BCT 18-MAR-2001  
 DEFINITION C.glutamicum lysE and lysG genes.  
 ACCESSION X96471  
 VERSION X96471.1 GI:1729753  
 KEYWORDS lysE gene; lysG gene; Lysine export regulator protein; Lysine exporter protein; Lysine governor.  
 SOURCE  
 ORGANISM  
 Corynebacterium glutamicum.  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
 Corynebacterium.  
 REFERENCE  
 1 (bases 1 to 2374)  
 Vrljic, M., Sahm, H. and Eggeling, L.  
 A new type of transporter with a new type of cellular function:  
 L-lysine export from Corynebacterium glutamicum  
 Mol. Microbiol. 22 (5), 815-826 (1996)  
 97126810  
 2 (bases 1 to 2374)  
 Vrljic, M.M.  
 Direct Submission  
 Submitted (07-MAR-1996) M.M. Vrljic, Institut fuer Biotechnologie  
 1, Forschungszentrum Juelich, Postfach 1913, D-52425 Juelich, FRG  
 FEATURES  
 source  
 1..2374  
 /organism="Corynebacterium glutamicum"  
 /strain="R127"  
 /db\_xref="taxon:1718"  
 complement(82..954)  
 /genes="lysG"  
 /complement(82..954)  
 /gene="lysG"  
 /note="Lysine governor"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Lysine export regulator protein"  
 /protein\_id="CAA65323.1"  
 /db\_xref="GI:1729754"  
 /db\_xref="SWISS-PROT:P94632"  
 /translation="MNPIQLDTLLSIDESFEGSLALSISPSAVSQRVKALEHHVG  
 RLVLSRTQPAKEAGEVILVQAAKMWLLQAEKKAOLSGRLAEIPILTIAINADSLSTW  
 FPFVNEVSGWAGTTLRLDEAHTLSLLLRGDVLGAVTREANPVAGCEVVELGTMK  
 HLAITPSLRDAYWDGKLDAAMPVLIRGPKDVLQDRDLGRVDGPNVGRRRYSVPS  
 AEGFGAIRLGLGWGLLPETQAAPMLKAGEVILLDEIPIDTPMIWQRWRLESRLARL  
 TDAVDAALTEGLRP"  
 1025..1726  
 /genes="lysE"  
 1025..1726  
 /gene="lysE"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Lysine exporter protein"  
 /protein\_id="CAA65324.2"  
 /db\_xref="GI:13397387"  
 /db\_xref="SWISS-PROT:P94633"  
 /translation="MEFITGLLLGSLILLSICPQNVLYIKQIKREGLIIVLVLLI  
 SDVFLFIAGTLVDLSNAPIVLDIMRWGGIAYLLWFAVMAAKDAMTKVEAPQIE  
 ETEPTVDFDPLGGSAVDTRNRVREVSVDKQVRWVPMALVILTLNPNAYLDA  
 FVFIGVGAQYGDTRWIFAAFGAFAASLIWFPLVGFCAALSRPLSPKVMRWINVV  
 AVVMTALAIKMLMG"  
 complement(1723..2352)  
 /genes="orf3"  
 complement(1723..2352)  
 /gene="orf3"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA65325.1"  
 /db\_xref="GI:1729756"  
 /db\_xref="SPRMBL:P94634"  
 /translation="MYALRDIYASGKALYVIGSIYSGPELTAAEAEMAEEGCPILLHQ  
 PSYIINRWVEPDPDGNLQSAANNGLVAFPSLAQGLLTDKYLIDGIPGSRASQ  
 GKSLSGMLNNVNDIMVRKINDIAQERGOSLAQMALAWVLREOGEYGAQTVTSALIGA  
 SSVEQLDNLSDLSLNLEFSDAELEAIDEISHDAGINIAKATDSKTRFN"

RESULT 3

A93933/c

LOCUS

DEFINITION

A93933 2374 bp DNA

Sequence 2 from Patent WO9723597.

PAT

22-JAN-2000

BASE COUNT 526 a 640 c 648 g 560 t

ORIGIN

Query Match 100.0%; Score 873; DB 2: Length 2374;  
 Best Local Similarity 100.0%; Pred. No. 1.le-235;  
 Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgaacccattcaactggacactttgtctcattatgatgaagcagcttgaagc 60  
 |||||  
 Db 954 ATGAACCCATTCAACTGGACACTTTGCTCTCAATCATGATGAAGCAGCTTGAAGC 895

Qy 61 gctcttagccctttccatttccctcgggtgagtcagcgcgttaagctctcag 120  
 |||||  
 Db 894 GCCTCCTTAGCCCTTTCCATTTCCCTCGGCGGTAGTCAGCCGTTAAAGCTCTCGAG 835

Qy 121 catcagtggtgagtggtgtatcgccacccacccgcaaacgcaaacgagcggt 180  
 |||||  
 Db 834 CATCAGTGGGTGAGTGTGGTATCGCGCACCCCAACCGGCCAAAGCAACGGAAGCGGGT 775

Qy 181 gaagtccttgtcaagcagcgcggaataatggtgtgctgcaagcagaactaaagc 240  
 |||||  
 Db 774 GAAGTCCTTGTGCAAGCAGCGCGGAAAAATGCTGTCTGCAAGCAGAACTAAAGCGCAA 715

Qy 241 ctatctgagcgcttgcgtaaatcccgcttaaccatcccatcaacgagagctctacc 300  
 |||||  
 Db 714 CTATCTGGACGCTTGTGTAATCCCGTTAACCATGCCCATCAACGGCAGATTCGCTATCC 655

Qy 301 acatggttctccctggttcaacgaggtagcttctgggtgagcagcgtcacctg 360  
 |||||  
 Db 654 ACATGGTTCTCCCGTGTTCACGAGGTAGCTTCTTGGGGTGGGCAACGCTCACGCTG 595

Qy 361 cgcttgaagatgaagcgacacattatccttgcctgcggcgtgagatgttttaggagcg 420  
 |||||  
 Db 594 CGCTTGAAGATGAAGCGCACACATTATCTTCTGCTGGCGGTGGAGATGTTTTAGGAGCG 535

Qy 421 gtaaccgctgaagctaataccgtggtggatgtgaagtagtagaactggaacatcg 480  
 |||||  
 Db 534 GTAAACCGTGAAGCTAATCCGTCGGCGGATGTGAATAGTAGAATGTGAACCATGCGC 475

Qy 481 cacttgccattgcaacccctcattcggtggtacacatggttgatgggaaactagat 540  
 |||||  
 Db 474 CACTTGCCATTGCAACCCCTCATTCGGGGATGCCCTACATGTTGATGGAACTAGAT 415

Qy 541 tgggtcgatgcccgtcttaccgttcggttcccaaaagtgtgcttcaagaccgtgacctg 600  
 |||||  
 Db 414 TGGCTCGGATGCCGCTCTTACGCTTCGGTCCCAAGATGTGCTTCAAGACCGGTGACCTG 355

Qy 601 gacggcgctcgatggtctctggtggcgagcgctcattcattgctccgtcgcgagaa 660  
 |||||  
 Db 354 GACGGCGCTCGATGTCTGTGGGCGGCGGCGGTATCCATGTCCCGTCGGCGGAA 295

Qy 661 ggtttggtgagcgaattcccgagggccttgggtgggactcttcccgaaacccaagct 720  
 |||||  
 Db 294 GGTTTGTTGAGGCAATTCGCCGAGGCGCTTGGTTGGGACTTCTTCCGAAACCAAGCT 235

Qy 721 gctccatgtctaaagcagagagtgatcctcctcagatagatacccatgacacaccg 780  
 |||||  
 Db 234 GCTCCCATGTAAAGCAGGAGAGTATCCTCCTCGATGAGATACCCTATGACACACCG 175

Qy 781 atgtattgcaagcagtgccctggaatctagatctcttagactcacagacgcgtc 840  
 |||||  
 Db 174 ATGTATTGGCAACGATGGCGCTGGAAATCTAGATCTCTAGCTAGACTACAGACCGCTC 115

Qy 841 gttgatcagaatcgagggttgcgaccttag 873  
 |||||

Db 114 GTTGATGCAGCAATCGAGGGATTGCGGCGCTTAG 82

ACCESSION A93933  
VERSION A93933.1 GI:6742037  
KEYWORDS  
SOURCE Corynebacterium glutamicum.  
ORGANISM Corynebacterium glutamicum.  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
Corynebacterium.  
REFERENCE 1 (bases 1 to 2374)  
AUTHORS Vrljic.M. and Eggeing.L.  
TITLE PROCESS FOR THE MICROBIAL PRODUCTION OF AMINO ACIDS BY BOOSTED  
ACTIVITY OF EXPORT CARRIERS  
JOURNAL Patent: WO 9723597-A 2 03-JUL-1997;  
KERNFORSCHUNGSANLAGE JUELICH (DE); VRLJIC MARINA (DE)  
FEATURES  
Location/Qualifiers  
1..2374  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
BASE COUNT 526 a 640 c 648 g 560 t  
ORIGIN

Query Match 100.0%; Score 873; DB 9; Length 2374;  
Best Local Similarity 100.0%; Pred. No. 1.1e-235;  
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaacccattcaactgacacattgtctcaatcattgatgaagcagcttcgaagc 60  
|||||  
DB 954 ATGAACCCATCACTGACACTTGTCTCAATCATTTGATGAAGCAGCTTCGAGGC 895  
|||||  
QY 61 gctcttagcctttccatttccctcctggcggtgagtcagcggtttaaactctcgag 120  
|||||  
DB 894 GCCTCTTAGCCCTTCCATTTCCCTCCCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGAG 835  
|||||  
QY 121 catcagtggttcgagtggttatcgccaccaccgcccgaagcaaacgagcagcggt 180  
|||||  
DB 834 CATCAGTGGGTGCGAGTGTGTGATCGCGCACCCCAACCGGCCCAACCGCGGT 775  
|||||  
QY 181 gaagtccttgcgaagcagcgagaaatggtgtctcaagcagcaaaactaaagcga 240  
|||||  
DB 774 GAAGTCTTGTGCAACGACGCGGGAATGTTGCTGCTCAGCAGCAAACTAAAGCGCAA 715  
|||||  
QY 241 ctatctggagccttgcgtgaatcccggttaacatgccatcccaagcagatcgctatcc 300  
|||||  
DB 714 CTATCTGGAGCGCTTCTGCTGAATCCGTTAAACATCGCCATCAACGACATTCGCTATCC 655  
|||||  
QY 301 acatggtttccctccgtgttcaacgaggtagctttcttgggggtggagcaagcgtcagctg 360  
|||||  
DB 654 ACATGTTTTCTCCCGGTGTCAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCGTG 595  
|||||  
QY 361 cgcttggagatgaagcgcacacattatcttgcgtcgcggtggagatgttttagagcg 420  
|||||  
DB 594 CGCTTGGAGATGAAGCGCACACATTATCTTCTGCTGCGCGGTGGAGATGTTTGGAGCG 535  
|||||  
QY 421 gtaacccgtgaagcctaactcccggtggcggtgtagtagtagtaagacttggaaaccatggcg 480  
|||||  
DB 534 GTAACCCGTGAAGCTAATCCGTTGGCGGATGTGAAGTAGTAGAACTTGAACCATCGCG 475  
|||||  
QY 481 caattggccattgcaacccctcattgcggagatgctacatggttgatggaaactaagt 540  
|||||  
DB 474 CACTTGGCCATTGCAACCCCTCATTTGCCGGATGCTTACATGTTGATGGGAACATAGAT 415  
|||||  
QY 541 tgggctgcgactcccgctttacgcttcgctcccaagaagtgtcttcaagacccgtgacctg 600  
|||||  
DB 414 TGGGCTGCGATGCCGCTCTTACGCTTCGCTCCCAAGATGTGCTTCAAGACCGTGAACCTG 355  
|||||  
QY 601 gacggcgcgctcgatggttccttggggcgacggcgcgatccattgtcccgctcgcgga 660  
|||||  
DB 354 GACGGCGCGCTCGATGCTCTTGGGGCGCAGCGCGGTATTCATTGTCCCGTCGCGGAA 295  
|||||  
QY 661 ggtttggtgagcgaattccgcagcgcttgggtgggactcttccggaaccccaagct 720  
|||||  
DB 294 GGTTTGTTGAGGCAATTCGCCGAGCGCTTGGTGGGAGCTTCTTCCCGAAACCAAGCT 235  
|||||

QY 721 gctcccatgctaaaagcaggaagtgatcctcctcgatgagatccccattgacacacg 780  
|||||  
DB 234 GCTCCCATGCTAAAGCAGGAGAGTGATCTCTCGATGAGATACCCATTGACACACG 175  
|||||  
QY 781 atgtattggcaacgatgcgctggaatagatctctagtagtagtcaacagcccgctc 840  
|||||  
DB 174 ATGTATTGCAACGATGCGCCTGGAATCTAGTCTCTAGTAGACTACAGACCGCGTC 115  
|||||  
QY 841 gttgatgcagcaatcgaggattcgccgcttag 873  
|||||  
DB 114 GTTGATGACCAATCGAGGATTGCGGCTTAG 82  
|||||

## RESULT 4

SC5F8/C SC5F8 33285 bp DNA BCT 09-JUN-2000  
LOCUS Streptomyces coelicolor cosmid 5F8.  
DEFINITION AL357613  
ACCESSION AL357613.1 GI:8347023  
VERSION  
KEYWORDS amino acid decarboxylase; anti-sigma factor; integral membrane protein; LuxR-family regulator; LysR-family transcriptional regulator; membrane transport protein; oxidoreductase; regulatory protein; RNA polymerase sigma factor; serine/threonine protein kinase; stress-inducible protein; TetR-family transcriptional regulator; thiorodoxin reductase; threonine dehydratase; two-component sensor histidine kinase.  
SOURCE Streptomyces coelicolor A3(2).  
ORGANISM Streptomyces coelicolor A3(2).  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 33285)  
AUTHORS Redenbach, M., Kieser, H.M., Denapalte, D., Eichner, A., Cullum, J., Kinashi, H. and Hopwood, D.A.  
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
Mol. Microbiol. 21 (1), 77-96 (1996)  
97000351  
JOURNAL 2 (bases 1 to 33285)  
MEDLINE Seeger, K.J. and Harris, D.  
REFERENCE 3 (bases 1 to 33285)  
JOURNAL Unpublished  
AUTHORS Thomson, N.R., Parkhill, J., Barrell, B.G. and Rajandream, M.A.  
TITLE Submitted (05-JUN-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
JOURNAL CB10 ISA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

## COMMENT

Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))  
CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream



ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we arrange for a overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid 5F8 lies on genomic restriction fragment AseI-A bordered by cosmids 5H1 and 4G10.

## FEATURES

source	1. .33285 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid 5F8" complement(1. .344) /gene="SC5F8.01c" complement(1. .108) /gene="SC5F8.01c" /note="nominal overlap with cosmid SC5H1 between bases 1. .108."
misc_feature	
CDS	complement(<1. .344) /gene="SC5F8.01c" /note="SC5F8.01c, possible serine/threonine protein kinase, partials CDS, len> 114 aa. Similar to many proposed Streptomyces coelicolor serine/threonine protein kinases e.g. TR:CA882014 (EMBL:AL161755) SCD63.07 (717 aa), fasta scores opt: 172 z-score: 222.0 E(): 6.8e-05 33.0% identity in 103 aa overlap. Overlaps and extends into CDS SC5H1.01 on the adjoining cosmid."
misc_feature	/codon_start=1 /transl_table=11 /product="putative serine/threonine protein kinase (fragment)" /protein_id="CAB93729.1" /db_xref="GI:8347024"
gene	/translation="MRAHVSTHELVAGRYRLFEVQVQTRNRCVSGEDATGTRGPLYVRLPEGRAGEAARAPGRVINTGETMASLCPGRIAPLDVAADGMLWTVTWEVAGVPLGLDLDRGA"
misc_feature	complement(492. .1451) /gene="SC5F8.02c" complement(492. .1451) /note="SC5F8.02c, possible threonine dehydratase, len: 319 aa. Similar to many including: Escherichia coli SW:THD2_ECOLI (EMBL:M21312) threonine dehydratase, catalytic (EC 4.2.1.16) TdCB (329 aa), fasta scores opt: 609 z-score: 676.6 E(): 3.2e-30 39.4% identity in 312 aa overlap. Contains a Prosite hit to Serine/threonine dehydratases pyridoxal-phosphate attachment site and a Pfam match to entry PF00291 S_T_dehydratase, Pyridoxal-phosphate dependent enzyme."
misc_feature	/codon_start=1 /transl_table=11 /product="putative threonine dehydratase." /protein_id="CAB93730.1" /db_xref="GI:8347025"
misc_feature	/translation="MTAITEIAAERAGIAGVVRVTPSPGLSALLGVPVTAKEELLQRTGSPKAGATAKLLSLEAEAGVAVSGNGHGIADVMAALDVKATVVMPTTA PARSVIAEAGALVRLTDGMSAFALVRLREGILVHPFDDPVPVAGQGVGLEF AEDASDLTVLSIGGGIAGVAAALRALRPGVRVGVWETEGEAMSRALAGGPUT VLVSSVVTLSVSRVSRVSLYVAELVTLVVPDREAVRGSLSALAEHAKVWTEPAAG CLLPAAVRVVERGDIAGLVVCGGNATVGDMAVWADRFLR"
misc_feature	complement(546. .1412) /gene="SC5F8.02c"
misc_feature	/note="Pfam match to entry PF00291 S_T_dehydratase, Pyridoxal-phosphate dependent enzyme, score 210.10, E-value 3.3e-59"
misc_feature	complement(1287. .1328) /gene="SC5F8.02c"
misc_feature	/note="PS00165 Serine/threonine dehydratases pyridoxal-phosphate attachment site"
gene	complement(1477. .2154)
CDS	complement(2166. .3470) /gene="SC5F8.04c" complement(2166. .3470) /gene="SC5F8.04c" /note="SC5F8.04c, unknown, len: 434 aa. Similar to Streptomyces coelicolor TR:Q9X9W9 (EMBL:AL096743) putative transport associated protein, SCI7.21 (399 aa), fasta scores opt: 428 z-score: 449.2 E(): 1.5e-17 38.3% identity in 209 aa overlap."
misc_feature	/codon_start=1 /transl_table=11 /product="hypothetical protein SC5F8.04c." /protein_id="CAB93732.1" /db_xref="GI:8347027"
gene	/translation="MSDLVPGGNVPLPGGVSVPVPGGFDVSLVYTDGKVGGRDADF VFNTPAFCARLDDTLTVDFARLRGAARVTVAVGSDPTGGLALPSPVTLVTDARGTVAFTPARGRETVLLVDFYRRGEGKVRALGQYADGLAGLDFGVEVTD APEPATPEPLTPGADSDGFLNLSNARAAGSPAVRPDPLRSARAAHAAADGTA TLTSETDGVSVHVRVVSAGFALTVEHLVSGRPTPAEFVAYCLRAERTRLTHAATA FTHAGWACVTPGSGDGYTALWAVPLTPGLARTTAEVVGLTNREACAGLPAALAVD ARLTAQAASHADMVTRDEYSHDTPGGAPDWRAAAAGADRRSNGENIACGQSPRAEV VEGMNSPGHRANILEAGTHGVGLAGGAGRGAGTYWTQLGG"
misc_feature	complement(3562. .6528) /gene="SC5F8.05c" complement(3562. .6528) /note="SC5F8.05c, possible LuxR-family regulator, len: 988 aa. Similar to several other putative regulators from Streptomyces coelicolor including: TR:Q9XAN4 (EMBL:AL079355) SC4C6.06 (893 aa), fasta scores opt: 305 z-score: 328.7 E(): 7.7e-11 26.4% identity in 910 aa overlap and TR:Q92573 (EMBL:AL035569) SC8D9.18 (1091 aa), fasta scores opt: 857 z-score: 926.1 E(): 0.36.5% identity in 1077 aa overlap. Contains a Prosite hit to PS00017 ATP/GTP-binding site motif A (P-loop) Pfam match to entry PF00196 GerE, Bacterial regulatory proteins, luxR family."
misc_feature	/codon_start=1 /transl_table=11 /product="putative LuxR-family regulator." /protein_id="CAB93733.1" /db_xref="GI:8347028"
misc_feature	/translation="MGGVQPTSLTSPVGREDEARLTCVLERARAGEARAVLACDA GVGKRTLHEAAGRAAAGTTLTCHVDLGDVGLPYLPFTTEILGLVANDERFNAVLA GHVADRLGGGPDGDDAAPSRLLFEGVAALLTDAVPLLVLEDLHWADQSSR DILLRFLGLGVQRPAGGAPRLALFASYRADDDCHRRHPLRLPLAEVLRPGVERLE LRPLPDSVARIYRLRRRLRSTVTRIVERAEGNAFYAEELVATAPAGVPSGL ADVLLIRFEQLSETAQOVLRTAAVAGRRVGHDLRLDRAVGLPEELESALREAVERQLL

Query Match 16.1%; Score 140.6; DB 3; Length 33285;  
Best Local Similarity 50.9%; Pred. No. 6.6e-29;

Matches 444; Conservative 0; Mismatches 414; Indels 15; Gaps 4;			
QY 13	caactggacactttgctctcaatcattgatgaagcagcttcgaagcgcctcttagcc 72		
Db 18092	CAGGTGGCACCCTGCTGGCGCTGCTGGACAGGACCTTCACAGCGCGCGCGCGCC 18033		
QY 73	ctttccatttccccctcggcggtgagtcgagcgggttaagactctcagcatcactggtggt 132		
Db 18032	CTCAGCTGACGCGCTCCGCGGTGAGCAGCGGTGAAGCACTGGAGCAGCGCACCGGC 17973		
QY 133	cgagtgttgatctcgac 192		
Db 17972	CGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 17913		
QY 193	caagcagcgcggaataatggtgctgctgcaagcagaaactaaacgcaactatctggaacgc 252		
Db 17912	CGGCTGGCG 17853		
QY 253	cttctgctg-----aatccccgttaaccatcgccatcaacgagattcgcctatccacatgg 306		
Db 17852	GGGAGGGCGGAGCG 17793		
QY 307	tttctccctgctgcttaacagaggttagcttttgggtggagcaacgctcaacgctgctgctg 366		
Db 17792	TTCTGCTGCG 17733		
QY 367	gaagatgaagcgcacacattactctgctgctgctgctgctgctgctgctgctgctgctg 426		
Db 17732	GAGGAGGAGGCGCACAGCG 17673		
QY 427	cgtgaagctaataccccctggtggtggtggtggtggtggtggtggtggtggtggtggtggt 486		
Db 17672	TCGTCGCGCGGACCGGTGCGCGGCTGACGCGTCCGCGAGCGTGGCGGATGCGCTATCTC 17613		
QY 487	gccattgcaacccctcattggtgggtggtggtggtggtggtggtggtggtggtggtggtg 543		
Db 17612	CCCTGCG 17553		
QY 544	gctgcgactgcccctcttaagcttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc 603		
Db 17552	GCG 17493		
QY 604	ggcgcgctcgatggttctgctggtggtggtggtggtggtggtggtggtggtggtggtggt 663		
Db 17492	CGGCTCGGTCACGCTGCG 17436		
QY 664	tttgcgtgaggaataatcgagcagccttgcgttgcgttgcgttgcgttgcgttgcgttgc 723		
Db 17435	TTGCTGAGGCGGCTGCG 17376		
QY 724	cccatgtaaaagcaggaagtgatctctctctg----atgatacccatcttgacacacgcg 780		
Db 17375	CCCTGCTGCGCACCG 17316		
QY 781	atgtattgcaacgattggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 840		
Db 17315	CTGTACTGCGACAGTGGGAAGCTGCACTCCCGCGCGCTGCGCGCGCGCGCGCGCGCGCG 17256		
QY 841	gttgatgcagcaatcagaggtggtggtggtggtggtggtggtggtggtggtggtggtggt 873		
Db 17255	GTGACGGCG 17223		

RESULT 5  
 MBU34849  
 LOCUS  
 DEFINITION  
 Mb34849  
 accession  
 version  
 keywords  
 source  
 organism

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.	
REFERENCE AUTHORS TITLE	1 (bases 10567 to 11392) Oettinger,T. and Andersen,A.B. Cloning and B-cell-epitope mapping of MPT64 from Mycobacterium tuberculosis H37Rv Infect. Immun. 62 (5), 2058-2064 (1994)
JOURNAL MEDLINE REFERENCE	94222581 2 (bases 1 to 14844) Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K. Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis J. Bacteriol. 178 (5), 1274-1282 (1996)
JOURNAL MEDLINE REFERENCE	96200095 3 (bases 1 to 14844) Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K. Direct Submission Submitted (24-AUG-1995) Mark J. Hickey, Molecular Micro., Pathogenesis Corp., 201 Elliott Ave. W., Seattle, WA 98119, USA
JOURNAL TITLE	Location/Qualifiers 1..14844 /organism="Mycobacterium bovis" /db_xref="taxon:1765" 2864..2888 /note="one of two imperfect direct repeats located at or near the deletion breakpoints"
misc_difference	replace(2888..13674,"") /note="this interval is completely absent in the closely related, but avirulent strain Mycobacterium bovis BCG"
gene	10705..11391 /gene="mpt64" 10705..11391 /gene="mpt64" /note="immunogenic protein"
CDS	/codon_start=1 /transl_table=11 /product="MPT64" /protein_id="AAC44034.1" /db_xref="GI:1049225" /translation="MRKIFMLVAVLLCCSGVATAAPKTYCEELKGTDTGACQIQ MSDPAYNINISLPYPOKSLNENIAOTRKFSLAATSSPREAPYELNITSATYQS AIPRGTOAVLVKYNAGGTHPTTYKAFDMDQAIRPFIYDTLWQADTDPLPVFP IVQGLSKOTGQVSIAPNAGLDPNYQNFVNDGVIFFPNPGELLPEAGPTQVLV PRSAIDSMIA"
repeat_unit	13652..13675 /note="one of two imperfect direct repeats located at or near the deletion breakpoints"
BASE COUNT	2800 a 4682 c 4603 g 2759 t
ORIGIN	
Query Match	14.0%; Score 122.6; DB 3; Length 14844;
Best Local Similarity	49.9%; Pred. No. 8.4e-24;
Matches 426; Conservative 0; Mismatches 409; Indels 18; Gaps 4;	
QY 13	caactggacactttgctctcaatcattgatgaagcagcttcgaagcgcctcttagcc 72
Db 4859	CAGCTGGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4918
QY 73	ctttccatttccccctcggcggtgagtcgagcgggttaagactctcagcatcactggtggt 132
Db 4919	TACATGTGTCACCCCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4978
QY 133	cgagtgttgatcgccac 192
Db 4979	CAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5038
QY 193	caagcagcgcggaataatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 250
Db 5039	CGGTTGGCGCGCAACACAGCGTGTGCTGAGTCCGAGCGCTGCTGCTGCTGCTGCTGCTGCTG 5098
QY 251	gcttgcgtgtaa----atccggttaaccatcgccatcaacgagattcgtatccacatgg 306

Db 5099 GCGTCGCTGAAACGACCGCATCACCATTGCGTAAACCGCGATTCATCGCGACATGG 5158  
Qy 307 ttctcctcgtttcaacgaggtagcttcttggggtgagcaacgctcagcgtcgcttg 366  
Db 5159 TTTTTCGGCGGTTCGAGC-----GTCTCGCGACGTCCTGCTCGAGTTCGGATC 5209  
Qy 367 gaagatgaagcgcacacattatctctgctgcggtgagatgttttaggagcgtgaacc 426  
Db 5210 GAGGACAGGACCATTCGCGCGGCTGCTACGAGAGGTGTGGGATGGCGCGGTGACC 5269  
Qy 427 cgtgaagtaataccgctgcggtgagatgaactagatagaactgaacacgccaacttg 486  
Db 5270 ACCGAGCGGAACCGGTGCGCGGCTGCGGGTGCACCGGCTGGGTGAAATCGGTACCTA 5329  
Qy 487 gccattgcaacccctcattcggtggtggtggtggtggtggtggtggtggtggtggt 546  
Db 5330 CCAGTGGCCAGCAGCGCATTCGTCAGCGGCATCTATCCGACGGGTTTCACATGCGCGCG 5389  
Qy 547 gcatgccccttctacgcttgcgtcccaagaatgtgcttcaagaccgtgacctggaacgg 606  
Db 5390 GCGGCTAAAGCTCCGCTCACTGCGGTGGAATCTGACGATGGGCTGCAGGACATGTTGGTG 5449  
Qy 607 cg---cgctgagtcctgtggtggcgagcgcgatctcattgtcccgctcgcggaaggt 663  
Db 5450 CGTAAGGCTTTTCGCGGCCATCACCAGACCGAGCGACTTTGTCCCGACCCACAGAGGGC 5509  
Qy 664 ttggtgaggaattcgcgaggtggtggtggtggtggtggtggtggtggtggtggtggt 723  
Db 5510 TTCACCGCGCAGCGCGCGGCTGGATGGGATGGGATGTTCCCGAGAGACTGCGACCA 5569  
Qy 724 cccatgctaaagcagagaagtagtctctctcctcctcctcctcctcctcctcctcctc 783  
Db 5570 TCTCCGCTTGGCGATGATGCTGCTAGCGGTCTCGACATACACCTCGAGTCCCTCTC 5629  
Qy 784 tattggaacgagcgtgagatctagatctctagctagactcaacagacgcgtcgctt 843  
Db 5630 TATTGGCAATCTGGAACCTGGACGTCCGATCCGATCCGCGCGAATTACCGACACGCTGAGG 5689  
Qy 844 gatcagcaatcg 856  
Db 5690 GCGGCGCAACGG 5702

RESULT 6  
LOCUS I86263 15239 bp DNA PAT 10-JUN-1998  
DEFINITION Sequence 17 from patent US 5700683.  
ACCESSION I86263  
VERSION I86263.1 GI:3205981  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 15239)  
AUTHORS Stover,C.Kendall and Mahairas,G.G.  
TITLE Virulence-attenuating genetic deletions deleted from mycobacterium BCG  
JOURNAL Patent: US 5700683-A 17<sup>th</sup> 23-DEC-1997;  
FEATURES  
source Location/Qualifiers  
BASE COUNT 2878 a 4791 c 4729 g 2841 t  
ORIGIN

Query Match 14.0%; Score 122.6; DB 10; Length 15239;  
Best Local Similarity 49.9%; Pred. No. 8.4e-24;  
Matches 426; Conservative 0; Mismatches 409; Indels 18; Gaps 4;  
Qy 13 caactgacacttgcctcaatcattgatgaagcagcttcgaagcgcctccttagcc 72  
Db 5253 CAGTGGCCGATTTGGTGGCGGTGCTGAACTGGGAGCTTCGATGCGCGCGGAGCGC 5312

Qy 73 ctttccatttccccctcggcggtgagtcagcgcgtttaagctctctcagacatcagctgggt 132  
Db 5313 CTACATGTCACCCCTCGGCTGTAGTCAGCGCATCAAGTCTGTGGAGCAGCAGGTGCGC 5372  
Qy 133 cgagtgttggtatcgcgaccccaacccggaacgaacgaacgaacgaacgaacgaacgaac 192  
Db 5373 CAGGTGCTGGTGGTTCAGGGAAGCCATGTCGGCGAGCAGCCGAGGTATCCCGCTGTTCG 5432  
Qy 193 caagcagcgcggaataatgtgtgctgcaagcagaactaaagcgcaactatctgg--ac 250  
Db 5433 CGGTGGCGCGCAACACAGCGCTGCTCGAGTCCGAGGGCTCGCTGAAATGGGTGGCAAC 5492  
Qy 251 gccttgctgaa-----atcccgtaacatcgccatcaacgagatctcgtatccacatgg 306  
Db 5493 GCGTGCCTGAAACACGACCGGATCACCATTGCGGTAAACGCCGATTCATGGCGACATGG 5552  
Qy 307 ttctcctcgtttcaacgaggtagcttcttgggtggtggtggtggtggtggtggtggtggt 366  
Db 5553 TTTTTCGGCGGTTCGAGC-----GTCTCGCGACGTCCTGCTCGAGTTCGGATC 5603  
Qy 367 gaagatgaagcgcacacattatctctgctgcggtggtggtggtggtggtggtggtggtggt 426  
Db 5604 GAGGACAGGACCATTCGCGCGGCTGCTACGAGGAGGTGTGGCGATGGCGCGGTGACC 5663  
Qy 427 cgtgaagtaataccgctggtgggtggtggtggtggtggtggtggtggtggtggtggtggt 486  
Db 5664 ACCGAGCGGAACCGGTGCGGGCTGCGGGTGCACCGCTGGGTGAAATCGGTACCTA 5723  
Qy 487 gccattgcaacccctcattcggtggtggtggtggtggtggtggtggtggtggtggtggt 546  
Db 5724 CCAGTGGCCAGCAGGCCATTCGTCAGGCGCATCTATCCGAGGGGTTTCACATGCGCGCGC 5783  
Qy 547 gcatgccccttctacgcttcctcctcctcctcctcctcctcctcctcctcctcctcctc 606  
Db 5784 GCGGCTAAAGCTCCGCTCACTGGCTGGAATCGTGAGCATGGCTCGAGGACATGTTGGTG 5843  
Qy 607 cg---cgctgagtcctgtggtggcgagcgcgatcattgtccctcctcctcctcctcctc 663  
Db 5844 CGTAAGGCTTTTCGTCGCGCATCACCAGACCGCAGCATTGTTGCCGACACAGAGGGC 5903  
Qy 664 ttggtgagcaattcgcgagccttgggtggtggtggtggtggtggtggtggtggtggtggt 723  
Db 5904 TTCACCGCGCGCGCGCGCGGCTGGATGGGATGGGATGTTCCCGAGAACTGGCAGCA 5963  
Qy 724 cccatgctaaagcagagaagtagtctcctcctcctcctcctcctcctcctcctcctcctc 783  
Db 5964 TCTCCGCTTGGCGATGGATGCTGTCAGCGGTCTCGACATACACCTCGAGCTCCCTCTC 6023  
Qy 784 tattggaacgagcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 843  
Db 6024 TATTGGCAATGCTGGAACCTGGACAGTCCGATCCGCGCATTCGCGCGAATTACCGACAG 6083  
Qy 844 gatcagcaatcg 856  
Db 6084 GCGGCGCAACGG 6096

RESULT 7  
MTCY39/c 7  
LOCUS MTCY39 38500 bp DNA BCT 17-JUN-1998  
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 89/162.  
ACCESSION Z74025 AL123456  
VERSION Z74025.1 GI:3261586  
KEYWORDS  
SOURCE Mycobacterium tuberculosis.  
ORGANISM Mycobacterium tuberculosis  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
Mycobacterium; Mycobacterium tuberculosis complex.  
REFERENCE 1 (bases 1 to 38500)  
AUTHORS Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,  
Harris,D., Gordon,S.V., Eigmler,K., Gas,S., Barry III,C.E.,

Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., and Squares, S., Scares, R., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G.

**TITLE** Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence

**JOURNAL** Nature 393 (6685), 537-544 (1998)

**MEDLINE** 98295987

**REMARK** Erratum: [[published erratum appears in Nature 1998 Nov

12; 396(6707):190]]

2 (bases 1 to 38500)

**REFERENCE** Parkhill, J.

**TITLE** Direct Submission

**JOURNAL** Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
On Jun 27, 1998 this sequence version replaced gi:1403436.

**COMMENT** Notes:  
Details of *M. tuberculosis* sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.  
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.  
**CAUTION:** In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

#### FEATURES

**source** Location/Qualifiers  
1..38500  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
<1..38419  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
/clone="Y39"  
complement(91..777)  
/gene="mpt64"  
complement(91..777)  
/gene="mpt64"  
/note="Rv1980c, (MTCY39.39), len: 228. mpt64 identical to MPB6\_MYCBO P19996 immunogenic protein mpb64/mpt64 (228 aa). Similar to Rv3036c, (MTV012.51c)"  
/codon\_start=1  
/transl\_table=11  
/product="mpt64"  
/protein\_id="CAA98382.1"  
/db\_xref="GI:1403475"  
/db\_xref="SWISS-PROT:P19996"  
/translation="MRKIFMLVTVAVLLCCSGVATAAPKTYCEBKJGDTGQACQIQ MSDPAYNINISLPYPOKSLNTIAQTRDKFLASSTPREAPYELNITSATYQS A1PPTGQAVLVKYNAGGTHPTTYKAFDWDQYRKRPITFDTLWQADTDLPLVYFP IYQGLSKQTQGVSTAPNAGLDPNVYQNFVAVTNDGVIFFFNPGELLPEAAGPTQVLV PRSAIDSLMA"  
complement(787..791)  
/note="ggagg, potential rbs upstream of Rv1980c"  
complement(968..1936)  
/gene="nrdF"  
complement(968..1936)  
/gene="nrdF"  
/note="Rv1981c, (MTCY39.38), ribonucleotide reductase

small subunit, len: 322, highly similar to RIR4\_SALFY PI7424, ribonucleoside-diphosphate reductase (319 aa) FASTA scores, opt: 1402, z-score: 1716.0, E(): 0, (66.0% identity in 315 aa overlap); Also similar to Rv3048c, (MTV012.63c); contains PS00368 Ribonucleotide reductase small subunit signature"  
/codon\_start=1  
/transl\_table=11  
/product="nrdF"  
/protein\_id="CAA98381.1"  
/db\_xref="GI:1403474"  
/db\_xref="SWISS-PROT:Q10840"  
/translation="MTGKLVVERHAINNRLLDADKQVWERLTGNFWLPEKIPLSND LASWOTLSTEQOTTIRVETGLTLDTAQTAVGAMIDDAVTPHEEAVLTNNAMFES VHAKSYSTFSTCTKOIDDAFDMSEQNPIQKRAQIIVDYRGDALKRKASVML ESFLYSGFYLPYMWSSRGKLTNTADLRLIIRDEAVHGYYIGYKQCRGLADLTDAER ADHREYTCELHTLYANEIDYAHLDYDELGWTDVLPYMKRYNANKALNMGYQPAFDR DTCQNPVAVRAALDFCAGENHDFFSGSGSSYYMGTHQPTTDDWDF"  
/gene="nrdF"  
complement(1589..1636)  
/note="PS00368 Ribonucleotide reductase small subunit signature"  
complement(1946..1949)  
/note="agga, potential rbs upstream of Rv1981c"  
complement(2161..2580)  
/gene="Rv1982c"  
complement(2161..2580)  
/gene="Rv1982c"  
/note="Rv1982c, (MTCY39.37), len: 139. Similar to Mycobacterium tuberculosis 292772|MTY20H10\_5 (131 aa) FASTA scores, opt: 288 z-score: 388.5 E(): 4.1e-14; 40.2% identity in 127 aa overlap, also similar to Rv2759c, and Rv0609"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein Rv1982c"  
/protein\_id="CAA98401.1"  
/db\_xref="GI:1403473"  
/db\_xref="SWISS-PROT:Q10874"  
/translation="MIVTSAVLVQCGRPHATLVAAALAGAHSPVMSAPTVABCLI VLTARHGPVARTIFERLSEIGLSVSSFEAAHAAATRAFLYGKGRHRAALFGDCM TYATAQLGHQPLLVANGNDFPQDTLEFRGVVGVPGVA"  
2981..2985  
/note="ggagg, potential rbs upstream of Rv1983"  
2992..4668  
/gene="PE\_PGRS"  
2992..4668  
/gene="PE\_PGRS"  
/note="Rv1983, (MTCY39.36c), len: 558; Member of *M. tuberculosis* PE\_PGRS family. Contains PS00141 Eukaryotic and viral aspartyl proteases active site; similar to MTCY130.10c, (43.3% identity in 522 aa overlap)"  
/codon\_start=1  
/transl\_table=11  
/product="PE\_PGRS"  
/protein\_id="CAA98400.1"  
/db\_xref="GI:1403472"  
/db\_xref="SWISS-PROT:Q10873"  
/translation="MSFLVYVPEFLTSAADVENIGSTLRANAAAAASTTALAAGA DEVSAAVALFARGEOEYQVSAQASAFHQFVOTLNSASGSYAAAEATIASOLQTAO HDLLGVNAPRTETLLGRPLIGDGAPCTATSPNGAGGLLYGNGNGYNTASGVGGGA AGSLGILNGAGGAGGPNAGGNGWLLNGGIGPGGASGIPMGSGAGTGGAAGLWGANGAGGAGLDGVGVDRGTGGAGGGLLYGGYGVSGPGGRTVPLBIH VTEPTVHANVAGGPTSTLLVDTSAGLVSPEDVGLVLMGLMGLTGLSISYSGGL YVIFATYTTVDFGNGIVTAPTAVNVLSTISPTSPFALSTVFALLADPTTTPFEAYF GAVGVGVLGVGNVAGPGPSIPTMALPDNLQGVLDAPAGELVFGNPLPAPNVEV VGSPTTLVVKIDGCTPIPVPSIIDSGVGTGTPISYVIGSGTLPTANTVIEVTPSGGD RLIAFNTNDYRPTVISSGLMNTGFLPFRFPQVYIDYSPSGIGTIVFDHPA"  
3871..3906  
/gene="PE\_PGRS"  
/note="PS00141 Eukaryotic and viral aspartyl proteases active site"  
complement(4656..5309)  
**misc\_feature**  
3871..3906  
/gene="PE\_PGRS"  
/note="PS00141 Eukaryotic and viral aspartyl proteases active site"  
complement(4656..5309)  
**gene**

```

CDS
/ gene="Rv1984c"
/ complement(4656) . 5309)
/ note="Rv1984c, (MTCY39.35), len: 217. Probable cutinase precursor with N-terminal signal sequence, similarity to CUT1, ALTBFR P41744 cutinase precursor, (209 aa) FASTA scores: opt: 283, z-score: 322.0, E(): 2.2e-11, (32.6% identity in 193 aa overlap). Also similar to M. tuberculosis Rv3452, Rv3451, Rv2301, Rv1758, Rv3724."
/ codon_start=1
/ transi_table=11
/ product="hypothetical protein Rv1984c"
/ protein_id="CAA98399.1"
/ db_xref="GI:1403471"
/ transi_table="MTPRSLVRIIVGVVAVTTALVSPAGRAAHADPCSDIAVVFAR GTHOAGSLGVGEAFVDSLTQVGVGVVAVNYPASDDYRASASNGSDASAHIOR TVASCPNTRIVLGGYSOGATVIDLSTAMPAPADHVAALFGEPSGSSMLWGG SLPTIGLYSSKINILCAPDPICTGGNIMAHVSYVQSGWTSQATFAANRLDHAG"
/ complement(4950) . 4967)
/ gene="Rv1984c"
/ note="PS00155 Cutinase, serine active site"
/ complement(5739) . 6650)
/ gene="Rv1985c"
/ complement(5739) . 6650)
/ gene="Rv1985c"
/ note="Rv1985c, (MTCY39.34), len: 303, LysR family member similar to many regulatory proteins, especially ICIA, ECOLI P24194, chromosome initiation inhibitor (297 aa), and P94632 LYSINE EXPORT REGULATOR PROTEIN. (290 aa). FASTA scores, P94632 LYSINE EXPORT REGULATOR PROTEIN opt: 705 z-score: 850.6 E(): 0; 42.7% identity in 288 aa overlap; and ICIA, ECOLI P24194 opt: 520, z-score: 632.5, E(): 1.1e-28, (35.8% identity in 285 aa overlap); contains PS00044 Bacterial regulatory proteins, lysR family signature. Also contains helix-turn-helix motif at aa 22-43, (+5.52 SD)."
/ codon_start=1
/ transi_table=11
/ product="hypothetical protein Rv1985c"
/ protein_id="CAA98410.1"

Query Match 13.9%; Score 121; DB 3; Length 38500;
Best Local Similarity 49.8%; Pred. No. 2.2e-23;
Matches 425; Conservative 0; Mismatches 410; Indels 18; Gaps 4;

Qy 13 caactgacatttgcctcaatcattgatgaagcagcttcgaagcgccctccttagcc 72
Db 6623 CAGCTGGCCGATGGCTGGCTGGTGAAGTGGCAGCTTCGATGGCGCCGCGAGCGC 6564

Qy 73 ctttccattccccctcgccggtgagtcagcgcgcttaagctctcgagcatcacgtgggt 132
Db 6563 CTACATGTCACCCGCTCGCTGTCAGTCAGCGCATCAAGTCGTTGGAGCAGCAGTCCGC 6504

Qy 133 cgagtggttgatcgccaccaccacccgcccgaagcagcagcggtgagtgctcttctgtg 192
Db 6503 CAGGTGCTGGTGGTCAGGGAAGCCATGTGCGGCGCAGCACCGCAGGTATCCCGCTGTG 6444

Qy 193 caagcagcgcggaataatggttctgcaagcagaactaaagcgaactatctgg--ac 250
Db 6443 CGGTGGCCGCGCAACAGACGCTGCTCAGTCGCGAGGGCTCGCTGAATGGGTGGCAAC 6384

Qy 251 gccttgctgaa----atcccgtaaccatgccatcacagcagattcgctatccacatgg 306
Db 6383 CGCTCGTGAACACGACGCGGATCACCATTGCGGTAAACGCGCATTCATCGCGCAGCATGG 6324

Qy 307 ttctccctcggttcaacgaggtagctcttgggtgggtggagcaacgctcacgctgcgttg 366
Db 6323 TTTTGGCCGCTGTCGACG-----GTCCTCGGCGACGCTCTGCTCGAGCTTCGGATC 6273

Qy 367 gaagatgaagcgacacattatctctgctgcgcgctgagatgttttaggagcgcgtaacc 426
Db 6272 GAGGACGAGGACCATTCGCGCGCGCTGCTACGGGAGGGTGTGGCGATGGCGCGCTGACC 6213

```

```

Qy 427 cgtgaagctaatacccgctggcggtggtgaagttagtagtaagaacctggaacctgcgcacttg 486
Db 6212 ACCGAGCGGAACCCGGTCCCGGGTGCACCCGGTGGGTGAAATCGCTACCTA 6153

Qy 487 gccattgcaacccctcattgcggatgcctacatggttgatggaaactagattggct 546
Db 6152 CCAGTGGCGAGCAGGCCCATTCGTCAGCGCCATCTATCCGACGGGTTCACCTGCCGCCGCG 6093

Qy 547 gcgatgcgcgtcttacgcttcacgcttcacaaagatgcttcaagacgcgtgacctgacggg 606
Db 6092 GCGGTAAAGCTCCGTCACCTGCGTGGAACTGTCAGCATGGCTGCAGGACATGTTGCTG 6033

Qy 607 cg---cgtcgatgctctgtgtggcgcgagcgcgctatccattgtccccctcgcggaaagt 663
Db 6032 CGTAAGGCTTTGCTGCGGCCATCACCAGACCGACGCACTTTGTGCCGACACAGAGGC 5973

Qy 664 ttggtgaggaattcgccgagcgtgtgttgggaactcttccccaaacccaagctgct 723
Db 5972 TTACCCGCGCAGCGCGCGCGGCTGGGATGGGCGATGTTCCCGGAGAACTGGCAGCA 5913

Qy 724 cccatgctaaagcaggaagatgctcctcgatgagataccattgacacacccgatg 783
Db 5912 TCTCGCTTGGCGATGGATCGTTCGTACGGGTCTGCGACATACACCTCGAGTCCCTCTC 5853

Qy 784 tattggaacatgagcgcgtggaatcttagatctcttagcttagctacacagacgcgtgctt 843
Db 5852 TATTGGCAATGCTGGAACACTGGACATCCGATCCGCGCAATATACCGACACGGTGAGG 5793

Qy 844 gatcgagcaatcg 856
Db 5792 GCGCGCGCAAGCG 5780

RESULT 8
AB011413/c 12070 bp DNA BCT 07-AUG-1998
LOCUS Streptomyces griseus genes for Orf2, Orf3, Orf4, Orf5, AfsA, Orf8,
DEFINITION partial and complete cds.
ACCESSION AB011413
VERSION AB011413.1 GI:3401946
KEYWORDS Orf8; AfsA; Orf5; Orf4; Orf3; Orf2.
SOURCE Streptomyces griseus
ORGANISM Streptomyces griseus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 12070)
AUTHORS Uneyama,T.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1998) to the DDBJ/EMBL/GenBank databases. Takashi Uneyama, University of Tokyo, Department of Agriculture and Life Sciences, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan (E-mail:aa67103@hongo.ecc.u-tokyo.ac.jp, Tel:+81-3-3812-2111)
2 (sites)
REFERENCE 2 (sites)
AUTHORS Uneyama,T.
TITLE Open reading frame encoded around afsA gene
JOURNAL Unpublished (1998)
FEATURES
source 1. 12070
Location/Qualifiers
/organism="Streptomyces griseus"
/db_xref="taxon:1911"
/complement(2860) . 3511)
/gene="orf2"
/complement(2660) . 3511)
/gene="orf2"
/function="sensor histidine kinase"
/codon_start=1
/transi_table=11
/product="Orf2"
/db_xref="GI:3401947"
/transi_table="VAAPVPAHPGPEAAAPAAARLTAAATLTITVWGLACLSWLIVIR
TDDRARMDAEYDEMRRRASVAASLVYVEDDIRSAGFAEHSILVLSIVIGAGALTALA

```

ATGTHVLSRSGCPANWALEQQERLIADAAHELRTPTVAVMRGVSVEVAAGAPGLEGQL  
 PRIRRAADRMADVNNLRTGRLEAGRTSYGRSLRLDLQDLVEVCAELPEGGPGLER  
 LEESVVAADAAALVRVAVNRLLNNAVRHGPGRGAAGRGRTAGDRVRTGVRVADGPVAV  
 APARLPS"  
 /function="Two-component regulator protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Orf3"  
 /protein\_id="BAA321131.1"  
 /db\_xref="GI:3401948"  
 /translation="MRVLVEEDDDLLRVIGAGLRDGGFAMDCASDMPADVLLHLLSA  
 YDCVLDKRVMPGSDTAPLEGRAGWSVPVLCTALDSLDRLEGRLESAGDYLAKP  
 FMRELVLVRGLSRALPSFLFCADVMDVARHEVRGGLVLSLSPKFAVAVLQO  
 LLVHRDVTYTRTGLLEHCWDEMADPVSNVVDVAVAGLRRLKGLSGPLVHTVRQGFLLS  
 ADGPGS"  
 complement(4333, >5250)  
 /function="methyltransferase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Orf4"  
 /protein\_id="BAA321132.1"  
 /db\_xref="GI:3401949"  
 /translation="NATQALGTALRLGVFDRICTELRAELAGSLGTHPRATLRLR  
 ALAGLQLLEPPEGAFRTTAAGNVLRADAPGTHVAMARFTDFVLRMGWDLDESVRT  
 GETTFDVGTFGLRHEHPELSAAFNEAMSGTRLTAEVTPHHYDFGRFORLVDIG  
 GGGTLLASILRAHOBEPGRVLEDTAEGLAQAAPRLAREGLDGRVLTLETGDFASAPAG  
 GDLYLNSIIHWDADAVRGLHRIHDVIPDHGSLLIIVELVLPATVPADRPDNNVLSDL  
 NMLNVNGRERTADDPAAALCTAGFCAGCA"  
 complement(5450, 6361)  
 /function="regulator protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Orf5"  
 /protein\_id="BAA321133.1"  
 /db\_xref="GI:3401950"  
 /translation="MSELPLDQVRLTLLAVVDEGTSDSATAALRLTPSAVSQRVKALEQ  
 RTGRVLLMTKTPVPTESGEVVALARQLARLEHAQAALMGSPGPEPTLLPIAVNAAQ  
 LATWSQPCGACRRWSATTAPGGPGVYRPAAGAGAGDGLVAGGGDGLFGFSAAG  
 RMRYLPVAPSGFADRWLRDGTALRELIEAPVVCVDRRDDLDQAFVRLRGLPGARPS  
 ARRLHVPVTEGSEFANAVASGMGMVPEVQAEPLLSDGRULVRLAPEPTVDVPLYWQWQK  
 LESPALAAVAEAAEAAEALDEAPPN"  
 7486..8391  
 /codon\_start=1  
 /transl\_table=11  
 /product="Afsa"  
 /protein\_id="BAA321134.1"  
 /db\_xref="GI:3401951"  
 /translation="WDAAEVVHPVGIEMVHRTPEDAFPNNVRLGRDRFAVEAVLP  
 HDHPFAPVGDLDHPLLVAEAMQAALAHAGYIGIPLGYHFLITELDYVCHPEHLG  
 VGGTEIGLEFVCSDLKWRAGLPAQGRVGMVHRGDRLAATGVAATRFSTPKAYRM  
 RGDVPEIGLSETPAPVPASPARARVEDVLSGTGREGVWELRVDTFRPTLQRPND  
 HYPGMLLEAARQAACLAVAGPAGIIPVEARTFRHYSEFGSPCWIGAVVQPCADETV  
 TVRVTHODGEVFEVSTVLSGPRANG"  
 10558..11598  
 /function="alcohol dehydrogenase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Orf8"  
 /protein\_id="BAA321135.1"  
 /db\_xref="GI:3401952"  
 /translation="MTTVAAYAPRAKAPLERPTTERRPVGFEFDILIDIKFAGICHSD  
 IHQANDGEGEIFFMPVGHETAGIVAETGPGVTRKVGDRVGVGMVDSGCTCDACL  
 GREGHACBENTQTYNALDRNGEPTTGYSTHLVTEKFAISPEGLADEAPLILCAG  
 ITTYSPLRWAGPKGVKVAAGVGLGHMAVKIAHALGAENTVLSQSLRKKDDGLKLG  
 ADHYATGDPRTFEELACTEDVILTSVAPLDQFAYLALLRTEGLVNVNVAPEEVS  
 NLFSLILNRSIAGSAIGIETQEMLDFCAVHGLCAIEVIGACQVNEATVERVLASD  
 VYRIVDTATI"  
 1647 a 4478 c 4272 g 1673 t

BASE COUNT  
ORIGIN

Query Match

12.4%; Score 108.6; DB 1; Length 12070;

Best Local Similarity 48.8%; Pred. No. 7.7e-20;  
 Matches 425; Conservative 0; Mismatches 424; Indels 22; Gaps 4;  
 QY 13 caactgagacattgtctctcaatcattgatgaaggcagcttctgaaggcgcctcttagcc 72  
 DB 6340 CAGGTTCGGACCTCTGGCGCTGGTGCAGCAAGGACCTTCGACTCGCGGGGACCGCC 6281  
 QY 73 ctttcattccccctcgcgcgtgagtcagcggttaaaagctctcagacatcagcggtg 132  
 DB 6280 CTGCGGCTGACCCCTGCGCGGTGTCAGTCACGGGTGAAGCGGTGGAGCAGGACCGGC 6321  
 QY 133 cgagtgttggtatcgcgcgcccccaaacggcggaagcaacacgagcggtgaagtccttgtg 192  
 DB 6220 CGGGTACTGCTGATGCGCAGAAAGCGGTGCGGCCACCGAGTCCGGTAGGTGGTGGCG 6161  
 QY 193 caagc---agcgcggaataatggtgtgctgcaagcagaagaactaaagcgaactatctga 249  
 DB 6160 CTCGCCCGCCAGCTGGCCAGGTGGAGCAGAGCGCGAGCGGCTGGGATGTCCGGG 6101  
 QY 250 cgcttctgtaaacctcccttaaccatcgccatcaacgcagattcgctatccacatggtt 309  
 DB 6100 CCGGGCAGCGCAGCTCTCTGCGGATCGCGGTGAACCTCCGATTCCTGCGGACCTGG--T 6043  
 QY 310 cctcccggttcaaacgagtagcttctt999tggagcaacgctcagcgtcgcttgtaa 369  
 DB 6042 CCTGGCAGCCCTCGCGCGGTGCGCGAGGAGTGGGACTCTGCTAGCAGCTGCGCGGGAG 5983  
 QY 370 gatgaagcgacacattatctctgctgcgcgtgagagatttttagagcggtaaacccgt 429  
 DB 5982 GACCAGGACCATACGCGCGCCCTGCTGCGGAGGGGTGCGGATGCGCGCGGTGACCTCG 5923  
 QY 430 gaagctaaccctggcggtgagtagtaagtagtagaacttgaaccatgcccacttgccc 489  
 DB 5922 TCGCCGAGGCGGTGACGGGCTGTCGGTCCGGCGGTGGCGCATGCGGTACTGCGCGG 5863  
 QY 490 attgcaacccctcattgaggagtgctacatggttgatgggaaactagattgggtgcg 549  
 DB 5862 TGCCAGTCCCGGTTTCGCGCAGCGGTGGGTGGGACGGCGGACCGCGCTTGGCGG 5803  
 QY 550 atg-----cccgcttctacgttcgctcccaagatgcttcaagacccgtg 595  
 DB 5802 AGCTGATCGCGGAAGTCCCGTGTCTGCTTCACAGGCGCGGACGATCTCCAGGAGCCT 5743  
 QY 596 acctgacggcgctcgatggttcctgtggcgcgagcgcgctatccattgtccctcgg 655  
 DB 5742 TCGTGGCGCGCTCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5683  
 QY 656 cggaaggtttggtgaggcaattcgccgagcccttggttggggacttctccgaaaccc 715  
 DB 5682 CGAGGGGTTCGCGAAGCGCGGTGGCTGCGGGGATGGGTGGGCGATGGTTCGCCGAGGTCC 5623  
 QY 716 aagctgctcccatgctaaaagcaggagaagtgtatcctcctcg---atgagatacccatg 772  
 DB 5622 AGGCGAGCGGTGCTGACGAGCGCGCGGTGCTGCGGGCTGCGCCCTGAGCCACCGTGG 5563  
 QY 773 acacaccgtatttggaacgagcgctcggaatctagatctctagtagtagtagtagtag 832  
 DB 5562 ACCTCCCGCTGACTGGCAGCAGTGAAGTGAATCGCCCGCGCGCGCGCGCGCGCGCG 5503  
 QY 833 acgcctgcttgatgacgaatacaggggatt 863  
 DB 5502 AGCGGTGGCGGAGCG 5472

RESULT 9  
 ECRPIAA/c  
 LOCUS ECRPIAA 1804 bp DNA  
 DEFINITION E.coli genes *iciA* and *rpiA*.  
 ACCESSION X73026  
 VERSION X73026.1 GI:405638  
 KEYWORDS ribose phosphate isomerase A.  
 SOURCE Escherichia coli.  
 ORGANISM Escherichia coli



Matches 223; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

```

QY 35 tcattgatgaagcagcttcgaagcgccctccttagccctttccattcccccctcgcg 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 693 TGATCGTGAACGAGGATTGACGCGCGGCACAAAAGCTGTCATTACAAATCAGCCG 752

QY 95 tgaagtcagcgcgttaagcttcgagcaccacgctgggtgagtggttgcgcaccc 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 753 TCTCAGCGCATTAAGCAACTGGAATAATATGTCGGCAGCGCGCTGTCGTCGATCCG 812

QY 155 aaccgccccaaagcaacgagcggtggaagtccttgcgaagcagcgagaaatggtg 214
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 813 TACCGCGCGCCCGACGAGGACAGGCAAAAGCTGTCGGCAGCTGTCGCGCAGTGAGT 872

QY 215 tgcgcagcagcaactaaagcgaactatctggaagccttgcgaaatcccgctaacca 274
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 873 TCGTGAAGAAGAGTGGCTGGCGGATGAACAAACCGGTTGCGACTCGCTCTTTCAC 932

QY 275 tgcgcataacgagattgcgtatcacatggttctcccggttcaacgaggtagctt 334
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 933 TGGCGGTCAACGCCGACAGCTGCGCGACGCTGTTGCTTCTGCACTGGCTCTGTTGG 992

QY 335 ctggggtgagcaacgctcacgctgcgttgaagataaagcgcacacattatccttc 394
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 993 CTGATGCGCTATCCGCTCAACTGACGTAGAGATGAACCCGCACTCAGGAACGTC 1052

QY 395 tgcggtgagatgttttaggagcggttaacccgtgaagcctaaccgtgcggtggtg 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1053 TGGCGCGCGGAGGTTGTCGCGCGGTGAGTATTCAACATCAGCGCTGCCGAGTTGTC 1112

QY 455 aagtagtaagcaggaacatgcgcacttggccat 491
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1113 TTGTCATAAACTTGTGGCGCTCGACTATCTGTTGCT 1149

```

```

RESULT 11
ECSEIRAICI ECSEIRAICI 6055 bp DNA BCT 08-JUL-1992
LOCUS E.coli serA, iclA, sbm genes and two open reading frames.
DEFINITION X66836
ACCESSION X66836
VERSION 1 GI:42941
KEYWORDS D-3-phosphoglycerate dehydrogenase; iclA gene; inhibitory factor;
          serA gene.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
          Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Escherichia.
REFERENCE 1 (bases 1 to 6055)
AUTHORS Roy,I.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1992) I. Roy, University of Cambridge, Dept of
          Biochemistry, CB2 1QW, UK
REFERENCE 2 (bases 1 to 6055)
AUTHORS Roy,I. and Leadlay,P.F.
JOURNAL Unpublished
FEATURES
          Location/Qualifiers
          1..6055
             /organism="Escherichia coli"
             /strain="K12-38"
             /db_xref="taxon:562"
             /map="62.8 minutes"
             /clone_lib="lambda EMBL three"
             /complement(1..298)
             /gene="serA"
             /complement(<1..298)
             /gene="serA"
             /EC_number="1.1.1.95"
             /note="partial of serA orf"
             /codon_start=1
             /transl_table=1
             /product="phosphoglycerate dehydrogenase"
             /protein_id="CAA47308.1"
             /db_xref="GI:42942"

```

```

/db_xref="SWISS-PROT:P08328"
/translation="MAKVSLEKDKIFLLVEGVHQALESRAAGYTNIEFHKGALDD
EOLKESIRDAHFILGRSRTHLTVEDVINAELKVAIGCFICGTNOVDLDAARARI"
complement(554..1249)
/note="orf1"
/codon_start=1
/transl_table=1
/protein_id="CAA47309.1"
/db_xref="GI:42943"
/db_xref="SWISS-PROT:P27252"
/translation="MRVKFHTTGETIMTQDELKKAAGVNAALQYVOPGTIVGVGTSTA
AFHIDALGTMKQIEGAVSSDASTEKLSLIGLHFDLNEVDSLGIYVDGADENGHM
QMIKGGAAITREKIIASVAERFICIAADAKQVLDLGRFPLPEVPIPARSAVADOLY
KLGRPEYRGVVTVDNGNVILDVHGMELDPIAMENAINAIPVVTGVLGFANRGADVA
LIGTPDGVQKTVK"
/1641..2534
/gene="iclA"
/1641..2534
/gene="iclA"
/note="iclA orf"
/codon_start=1
/transl_table=1
/product="inhibitor of chromosome initiation"
/protein_id="CAA47310.1"
/db_xref="GI:42944"
/db_xref="SWISS-PROT:P24194"
/translation="MKRPDYRTQALDAVIRERGFERRAQAQKLCITQSAVSQRIKOLEN
MFGQPLLVRTPPRTEQOKLALLROVELLEEWLGEOTGSTPLLLSLAVNADSL
ATWLLPALAPVLADSPIRNLQVDETRTQERLRGEVGVAVSIHQALPCLVLDKLG
ADWLLFVSKPFAEKYFPNGVTRSAKLLKAPVAVFDHDDMHQAFQONFDLPPGVPYC
HIVNSSEAFVQLARQCTTCMTPHQLQIEKELASGELIDLTPGLFORMRLYWHRFAPES
RMRKVTDALDLYGHKVLROD"
2737..4881
/note="sbm orf"
/codon_start=1
/transl_table=1
/protein_id="CAA47311.1"
/db_xref="GI:42945"
/db_xref="SWISS-PROT:P27253"
/translation="MSNVQEQWOLANKELSRREKTVDSLHOTAEGIAIKPLYTEADL
DNLEVTGLPLGPPVYPRATMYTAQPTROYAGFSTAKESAFYRNLAAAGOKGL
SVAFDLATHRGVDSNPRVAGDVAGVAIDTVDMKVLFDQIPLDKKSYSMTMGAV
LPVAFYVAAEEQVTPDKLTGTONDILKELCYRNTIYPPKFSMRIDIAIWCSS
GNMFRNTISISYHMGAGANCVOOAVTADIEYKAAISAGLKIDDFAPRUSFF
FGIMDLFMVAMLRARYLWSEAVSGCAQDPKSLARHTCQTSGLSELCRTVDPL
IRTTLEAATLGGTQSLTNVDFEALGLPTDSARIARNTQIIIESESELCTVDPL
AGSYVIESITDQIVKOARAIQOIDEAGMAKAEAGLPKRMIEASAREOSLIDQK
RVIVGVNKKLDHEDETQVLEIDNVVRNEQIASLERIRATPDAAVTAALNATHAA
QHNEILAAVNAARVRATLGEISDALEAFVFDLYVPSQCVTVTAQSTHQSEKASE
FDAIYAQTEQFLADNGRRPRILIAKMGQDGHDRGAKVIASYSDLGDFVDJSPMFSTP
BEIARLAVENDVHVVGASSLAAGHKTLLPELVALKKWGREDICVYVAGVIPPQDYAF
LQERGVAAIYGPCTPMLDSVRDVLNLSQHHD"
4948..5754
/note="orf2"
/codon_start=1
/transl_table=1
/protein_id="CAA47312.1"
/db_xref="GI:42946"
/db_xref="SWISS-PROT:P27254"
/translation="MTLVESRHPHQALSTQLLDAIMPYCGNTLRLGVGTGTGACKST
FLEAFGLLIREGLKVAIVADVPSPVTVGGSLGDKTRNDLARAFAAFIRVPSSGH
LGGASQARLEMLICEAAGYDVIVETVVGQSETEVARMVDCFLISLQIAGGGDLQK
IKKGLMEVADLTVINKDDGNHTNVAIARHVESALHILRRKYDEWQPRVLTCSALEK
RGIDEIHAIIDFKTALTASGRLOQVROQSQSVEMLRKQTRRRSTESPVE"

```

BASE COUNT 1477 a 1615 c 1550 g 1413 t

ORIGIN

Query Match 9.5%; Score 82.6; DB 2; Length 6055;

Best Local Similarity 48.8%; Pred. No. 1.8e-12;

Matches 223; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

QY 35 tcattgatgaagcagcttcgaagcgccctccttagccctttccattcccccctcgcg 94



Db 1684 TGATACGTGAACGAGGATTTGAGCGCGCGCACAAAAGCTGTGATTACAAATCAGCCG 1743  
 Qy 95 tgaagcagcgcttaagctctcagcatcacgtgggtcgagtggtgtatcgcgcaccc 154  
 Db 1744 TCTCACAGCGCATTAAGCAACTGGAATAATGTTCCGGCAGCCCGCTGTGTGGTACCG 1803  
 Qy 155 aaccgccccaaagcaacgaagcgggtgaagtcctgtgcaagcagcgcggaataatggtg 214  
 Db 1804 TACCGCGCGCGCCGACGGAACAGGCAAAATGCTGGCAGCTGTGCGCAGGTGGAGT 1863  
 Qy 215 tgcgtgaagcagaactaaagcgaactatctggaacgcttgcgtgaaatcccgtaaccca 274  
 Db 1864 TGCTGGAAGAAGAGTGCGTGGCGCATGAACAAACCGGTTGCACTCCGCTGCTGCTTTCAC 1923  
 Qy 275 tcgcatcaacgcagattcgcattcccatggtttccctcccggtttcaacgaggtagctt 334  
 Db 1924 TGGCGGTCAACGCCGACAGCTGCGGAGTGTTGCTTCCTGCAGCTGGCTCTGTGTGG 1983  
 Qy 335 cttggggtgagcaacgcctcagctgcgtggtggaagatgaagcgcacacattactctg 394  
 Db 1984 CTGATTGCGCTATCGGCTCAACTTGCAGGTAGAAGATGAACCCGCACTCAGGAACGTC 2043  
 Qy 395 tgcggtgagatgttttagagcggtaaacccgtgaagctaataccggtgcggagatgtg 454  
 Db 2044 TGCGCCGCGCGGAAGTGTGCGCGGTTGAGTATTCACATCAGCGCGCTGCCGAGTTGTC 2103  
 Qy 455 aagtagtagaactggaaccatgcccacttggccat 491  
 Db 2104 TTGTCGATAACTGTTGGCGCTCGCACTATCTGTTGCT 2140

## RESULT 12

AE000375 LOCUS AE000375 10362 bp DNA BCT 01-DEC-2000  
 DEFINITION Escherichia coli K12 MG1655 section 265 of 400 of the complete genome.  
 ACCESSION AE000375 U00096  
 VERSION AE000375.1 GI:1789282  
 KEYWORDS  
 SOURCE Escherichia coli K12.  
 ORGANISM Escherichia coli K12.  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.  
 REFERENCE 1 (bases 1 to 10362)  
 AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.  
 TITLE The complete genome sequence of Escherichia coli K-12  
 JOURNAL Science 277 (5331), 1453-1474 (1997)  
 MEDLINE 97426617  
 PUBMED 9278503  
 REFERENCE 2 (bases 1 to 10362)  
 AUTHORS Blattner, F.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459  
 REFERENCE 3 (bases 1 to 10362)  
 AUTHORS Blattner, F.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459  
 REFERENCE 4 (bases 1 to 10362)  
 AUTHORS Plunkett, G. III.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
 This sequence was determined by the E. coli Genome Project at the

University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). \*\*\* The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

## FEATURES

source Location/Qualifiers  
 1..10362  
 /organism="Escherichia coli K12"  
 /strain="K12"  
 /sub\_strain="MG1655"  
 /db\_xref="taxon:83333"  
 complement(9..40)  
 /note="factor Sigma70; predicted +1 start at 3057674"  
 34..63  
 /note="factor Sigma70; predicted +1 start at 3057742"  
 101..994  
 /gene="icla"  
 /note="b2916"  
 101..994  
 /gene="icla"  
 /function="regulator; DNA - replication, repair, restriction/modification"  
 /note="b2917; 100 pct identical to ICIA\_ECOLI SW: P24194; CG Site No. 28646"  
 /codon\_start=1  
 /transl\_table=11  
 /product="replication initiation inhibitor, binds to 13-mers at oriC"  
 /protein\_id="AAC75953.1"  
 /db\_xref="GI:1789283"  
 /translation="MKRPDYRTLQALDAVIRERGFRAQKLCITQSAVSQRIKQLEN MFGQPLIVTPPPRTTEQOKLLALLRQVELLEEWLGDEQGTSTPLLLSLAYNADSL ATWLLPALAPVADSPIRNLNLOVEDTTRTORRGEVGVGAVSIQHQALPSCIIVDKLG ALDYLVSXKPKFAEKYFPKGVGTSALLKAPVAFDLDMDHQAFLQGNFDPGVSVPCHIVNSSEAFVQLARQGTTCMIPHQLQIEKELASGELLDTPLGLFQRRMLYWHRAPEES RMRKVTDALLDYGHKVLROD"  
 1045..1072  
 /note="factor Sigma70; predicted +1 start at 3058751"  
 1198..3342  
 /gene="sbm"  
 /note="b2917"  
 1198..3342  
 /gene="sbm"  
 /function="enzyme; Degradation of small molecules: Fatty acids"  
 /note="o714; 99 pct identical to SBM\_ECOLI SW: P27253"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methylmalonyl-CoA mutase (MCM)"  
 /protein\_id="AAC75954.1"  
 /db\_xref="GI:1789284"  
 /translation="MSNVQBWQQLANKELSRREKTVDSLHVHQTAEGIAIKPLYTEADL

```
gene
DNLVETGLPLPPYVGRPRATMYTAQPTWIRQYAGFSTAKESNAYFYRNRLAAGOKGL
SVAFDLATHRGYDSNPRVAGDVGKAGVAIDTVEDMKVLFQDIPDKMSVSMTHGAV
LPVLAIFYAAEEQGVTPDKLTQAGNDILKLEYLRNTYIYPKPSMRITADIAGV
GNMPRENTISISYHMGAGAGCVOQAFETLADIEYKAAISAGLKIDDFAPLSFF
FGIGMDLNNVAMLAARAYLWSEAVSGFGAOPDKSLALRTHOOTSWSLTEODPNVY
IETTEALATLGCTQSLHTNAFDEALGLPTDFSARIARTQIIIOESELCTVDPL
AGSYTESDQIVKQARALIOOIDEAGGMAKAIKAGUPKRMIEESAREQLIDOGK
RVIVGNKYKLDHEDETVDLEIDNVMRNEQIASLERIRATRDDAAVTAALNALTHAA
QHNEILLAAVNAARVATPLGEISDALEAFDRYLVPSOCVTGVIQAQSHQEKASAE
FDAIVAOEFOFLADNGRRPRILIAKMGODHGRGAKVIAASVSDLGFDVDSLSPMESTP
EELARLAVENDVHVGCASSLAAGHKTLPPELVEALKKWGREDICVAVAGVIPPQYAF
LQERGVAAIYGCTPHLDSVRDVLNLSOHHB"
3335. .4330
/gene="y9fd"
/note="y9fd"
3335. .4330
/note="b2918"
3335. .4330
/gene="y9fd"
/function="phenotype; Not classified"
/note="0331; 100 pct identical to yGFD_ECOLI SW:
p27254(257 aa) but has 25 additional N-ter aa 50
additional C-ter aa; orf2 of x66836"
/codon_start=1
/transl_table=11
/product="putative nucleotide-binding protein"
/protein_id="AAC75955.1"
/db_xref="GI:1789285"
/translacion="MINATLAEISIRLRQGERATLAQAMTLVESRHRPHOALSTOLL
DAIMPCTNDLRGLVGTGPGAGKSTFLAEGMLLIREGLKAVIAVDPSPPVTGGSIL
GDKTMRNDLARAFAIRPVPSSGHLGASORARELMILCEAAGYDVIVETVGVGQS
ETEVARWDCFTSLQIAGGDDLOGIKKGLMEVADLIIVNKDDGNHNVAIARHME
SALHILREYDQWPRFVTCRSLERKRIDEIWHAIIDFKTALTALTSRQOQVROQVSE
WURKQTEEVNLHMFANEDFDYRQTLVAVKNNTLSRPTGLRQLSEFIQYQYFD"
4299. .5126
/gene="y9fg"
/note="b2919"
4299. .5126
/gene="y9fg"
/function="putative enzyme; Not classified"
/note="0275; This 275 aa ORF is 30 pct identical (6 gaps)
to 250 residues of an approx. 296 aa protein ECHM_RAT SW:
p14604"
/codon_start=1
/transl_table=11
/product="putative enzyme"
/protein_id="AAC75956.1"
/db_xref="GI:1789286"
/translacion="WNLGRNLIKGFMSYQVNVVITINKVAVIEENVGRKLNALSK
VFIDLMQALSDLNRPETIRILRAPSGSKVFSAGHDHELPSGGRDPLSYDDPLRQI
TRMIQKFKPILSVWEGSVWGAFEMINSDDLIIAASSTFSMTFVNLGVYVNLVGIH
NLTRDAGEHIVKELIFTASPIAQALAVGILNHVVEELEDFTLQMAHHSKAPL
AIAVKEELRVLGEARTWNSDEFERIQGMRRAVYDSYQEGMNAFLKRPNFVGH"
5150. .6628
/gene="y9fh"
/note="b2920"
5150. .6628
/gene="y9fh"
/function="putative enzyme; Not classified"
/note="0492; This 492 aa ORF is 46 pct identical (7 gaps)
to 490 residues of an approx. 544 aa protein CATL_CLOKL
SW: p38946"
/codon_start=1
/transl_table=11
/product="putative coenzyme A transferase"
/protein_id="AAC75957.1"
/db_xref="GI:1789287"
/translacion="METOWTRMTANAEAIIOHNDMAVFSGFTPAGSPKALPTATARR
ANEQEAQKPYOIRLLTCASISADADVLSADAVSWRAPYOTSGLRKKKNOGAVF
VDLHLSVAQWVNYGFGDIDVAVIEASALAPDGRVLTSGIGNAPTWLLRAKKVIE
LGHNDHPRVAELADIVIPGAPPRNSVSIFHAMDRVGRYVOIDPKKIVAVVETNLPD
AGNMIDQWCMQCOQADNVVTFLLQEMAHGRIPPEFLPQSGVGNINNAVARLGNP
VIPPNNPSEVLEQSVSHLLETGKISGASASLPTTSADSLRKIYDNMDYFASRIVLRP
OEISNNPTEIRRLGVIALNVGLEFDIYGHANSTHVAGVDLMNIGGSGDFERNVLSI
FMAPSIAREGSTVVPWPCSHVDHSEHSVKVIITEQGIADLRGLSPLOQARTIIDNCA
```

```
gene
HPMYRDYLRHYLENAPGGHIHHDLSHVFDLHRNLATGSMGLG"
complement(6625. .7536)
/gene="y9fi"
/note="b2921"
complement(6625. .7536)
/gene="y9fi"
/function="putative regulator; Not classified"
/note="f303; This 303 aa ORF is 25 pct identical (11 gaps)
to 241 residues of an approx. 304 aa protein CYNR_ECOLI
SW: p27111"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator LYSR-type"
/protein_id="AAC75958.1"
/db_xref="GI:1789288"
Query Match 9.5%; Score 82.6; DB 1; Length 10362;
Best Local Similarity 48.8%; Pred. No. 1.7e-12;
Matches 223; Conservative 0; Mismatches 234; Indels 0; Gaps 0;
QY 35 tcattgatgaagcagcttcgaagcgccctccttagcccttcattcccccctcgcg 94
DB 144 TGATACGTGAACGAGGATTGACGCGCGGCACAAAAGCTGTCATTACAAATCAGCCG 203
QY 95 tgagtcagcgcggttaaaagctctcgagcatcacgtgggtcgagtggttggtatcgccaccc 154
DB 204 TCTCACAGCGCATTAAGCAACTGGAAAAATATGTCGGGACGCGCTGTTGGTCGTACCG 263
QY 155 aaccggcgaagcaaacgaagcggtggaagtccttgcgaagcagcgcggaataatggt 214
DB 264 TACCGCGCGCGCGACGAGTGAAGCAAGGCAAAACTGCTGTCACATGCTGCGCAGGTGGAGT 323
QY 215 tgcgtcaagcagaactaaagcgcaactatctgagcgcttgcgaaatcccgqtaacca 274
DB 324 TGCTGGAAGAAGAGTGGCTGGCGGATGACAAACCGGTTGCACTCCGCTGCTTTTAC 383
QY 275 tgcgcataacgcagatctgcctatccacatggtttcctcccgtgttcaacgaggtagctt 334
DB 384 TGGCGGTCAACGCCGACAGTCTGGCGACGTGTTGCTGCTGTCACATGCTGCTGTTGG 443
QY 335 cttggggtgagcaagcgtcacgctgcgtgctggaagataagcgcaacattatccttc 394
DB 444 CTGATTGCGCTATPCCGCCTCAACTTCGAGGTAGAAGATGAACCCGCACTCAGGAACGTC 503
QY 395 tgcggcggtgagatgttttaggagcggttaaccgctgaagcctaaccgtggcggtggtg 454
DB 504 TCGCGCGCGCGGAAGTGGTGGCGCGGTGAGTATTCAACATCAGCGCGCTCGCGAGTTGTC 563
QY 455 aagtagtagaacttgggaacacatgcgccacttggccat 491
DB 564 TTGTCGATAAACTTGTGGCGCTCGACTATCTGTTCGT 600
RESULT 13
AE005521 10367 bp DNA BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 140
DEFINITION of 290.
ACCESSION AE005521 AE005174
VERSION AE005521.1 GI:12517435
KEYWORDS
SOURCE Escherichia coli O157:H7 EDL933.
ORGANISM Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (bases 1 to 10367)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
```

```

JOURNAL Nature 409 (6819), 529-533 (2001)
MEDLINE 21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 10367)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
    source
        Location/Qualifiers
            /organism="Escherichia coli O157:H7 EDL933"
            /strain="EDL933"
            /serotype="O157:H7"
            /db_xref="taxon:155864"
            /note="enterohemorrhagic"
            complement(357..1559)
            /gene="vsc"
            /notes="24243"
            complement(357..1559)
            /gene="vsc"
            /function="orf; Unknown function"
            /notes="Residues 1 to 400 of 400 are 98.75 pct identical to
            residues 1 to 400 of 400 from Escherichia coli K-12 Strain
            MG1655; B2906"
            /codon_start=1
            /transl_table=11
            /product="orf, hypothetical protein"
            /protein_id="AAG58033.1"
            /db_xref="GI:12517436"
            /translation="MOSVDVAIVGGMGVLAVAGLGGSLRVAVLEORVPEPLAADA
            PPOLRYSAINAASEKLLRLGLWODILSRASCYHGMEDKDSRGHIFSDGSMYCS
            HLGHIYNSVHYHNLNKAQOQSDITLLAPAELOQVANGENEFILKDGSMITARLV
            IGADGANSJURNKADIPLTWYOHHALVATIRTEPHDAVARQVFHGEIGLAFPLPS
            DPHLCISVLSLSPAEQMOQASEDEFNALNTAFONRLGLCKRVSARQVFPITGRYA
            ROPAAHRLALVGAHTIHPLAGQGVNLGFMADAEITAEKRLHRQCKDGIQYILRR
            YRSRKHSAALMAGMQGRDLFSGTNPVKLLRDIGLKLADTLPGVKQLIRQAMGL
            NDLPENLR"
            complement(1583..2761)
            /gene="ubiH"
            /notes="24244"
            complement(1583..2761)
            /gene="ubiH"
            /function="enzyme; Biosynthesis of cofactors, carriers:
            Menaquinone, ubiquinone"
            /notes="Residues 1 to 392 of 392 are 98.97 pct identical to
            residues 1 to 392 of 392 from Escherichia coli K-12 Strain
            MG1655; B2907"
            /codon_start=1
            /transl_table=11
            /product="2-octaprenyl-6-methoxyphenol-->
            2-octaprenyl-6-methoxy-1, 4-benzoquinone"
            /protein_id="AAG58034.1"
            /db_xref="GI:12517437"
            /translation="MSVIVGGMGAGATLALISRLSHGALPVHLLIETAPESHAPHG
            FGRALTAGTCQQLARIGVMSLADCATITTVSDRGHAGFVFLAEEDQLAAL
            GOVELHNVGQRFALLRKAPGYTLHCPDRVANVARTQSHVEVTLEGGTTLGRVLVA
            ADGTHSALATVCGDWMQOEPYEQLAIVANVATVAHEGRAFTQHGFLAMLPMDSG
            RCLVWCHPLERREELVSWDEKFCRELQSAFGWRLKITHAGKRSAYPLALTHAARS
            ITHRTVLVGAQOTLHPIAGOGFNLGMRDVMSLAETLTAQGERGDMGDYGLCRYQQ
            RQSDREATIGYVDSLVHLFANRWTFVLVGRNIGLTMELTFPARDVLAQRYXGWVAR
            "
            complement(2758..4083)
            /gene="pepp"
            /notes="24245"
            complement(2758..4083)
            /gene="pepp"
            /function="enzyme; Macromolecule degradation: Degradation
            of proteins, peptides, glyco"

/note="Residues 1 to 441 of 441 are 99.09 pct identical to
residues 1 to 441 of 441 from Escherichia coli K-12 Strain
MG1655; B2908"
/codon_start=1
/transl_table=11
/product="proline aminopeptidase P II"
/protein_id="AAG58035.1"
/db_xref="GI:12517438"
/translation="MSEISRQEFQRRRQALVEQMFGSAALIFAPEVTRSDSEYPY
RQNSDFWYTFGFNEPEAVLVLIKSDDTHHNSVLFNVRDLTAIEFGRRLGQDAAPK
ATVDALAFSEINQQLGLDGVVYHAQGEYAYADEIVNSALEKRLKRGSRONLTAP
LMDWRPVVHEMRLEKSPSEETAVLRAGEITAMATRAMKRCRCMEVHLEGETHH
EFNRHGRAPSYNTIVSGGNGCILHYTENECMRDGLVIDAGCEYKYGADIRRT
FPVNGKPTAQOREIYDIVLESUETSURLYRPGTSLIEVTEGVVRVMSGLVGLIKRG
DIDELIAQNAHRPFPMHGLSHWGLDGVVYQDGRSRILEPQWMLTVPEPGLYIAPD
ADVPEQYRGIGRIEDDITETGNTENLTASVVKKPEETEAALMAAARKQ"
complement(4109..4693)
/gene="ygfB"
/note="24246"
complement(4109..4693)
/gene="ygfB"
/function="orf; Unknown function"
/note="Residues 1 to 194 of 194 are 100.00 pct identical
to residues 1 to 194 of 194 from Escherichia coli K-12
Strain MG1655; B2909"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG58036.1"
/db_xref="GI:12517439"
/translation="MLMSIQNEMPGYNEMNOYLNQOCTGLTPAEMHGLISGMICGGND
DSMLPLLDLTNEGMAFGHQAQAKRMHSATSDALQDDGFLFQLYLPDGDVSVFD
RADLAGWVNHFLGLGVTPQKLDVKTGETGEAIDDLRNLIAQLGYDEDEDELEMSL
EETIEYVRVAALICHDTFTHPQTAPEVQKPTLH"
4855..5184
/gene="ygfE"
/note="24247"
4855..5184
/gene="ygfE"
/function="orf; Unknown function"
/note="Residues 1 to 109 of 109 are 100.00 pct identical
to residues 1 to 109 of 109 from Escherichia coli K-12
Strain MG1655; B2910"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG58037.1"
/db_xref="GI:12517440"
/translation="MSAQPDVQIFGRSLRVNCPDPORDALNQAADLNRLODLKER
TRVTNTEQVLVFTAAALNISYELAQEKARTRDYAASMEQIRMLQQTIEQALLBQGRITE
KTNQNP"
5225..5408
/gene="ssrS"
/product="6S RNA"
/function="RNA; Macromolecule synthesis, modification:
Ribosomal and stable RNAs"
5225..5408
/gene="ssrS"
/note="24248"
5484..6032
/gene="ygfA"
/note="24249"
5484..6032
/gene="ygfA"
/function="putative enzyme; Not classified"
/note="Residues 1 to 182 of 182 are 99.45 pct identical to
residues 1 to 182 of 182 from Escherichia coli K-12 Strain
MG1655; B2912"
/codon_start=1
/transl_table=11
/product="putative ligase"
/protein_id="AAG58038.1"
/db_xref="GI:12517441"

```

```

/translation="MIRORRRALTPQQQEMQQQAATRMNTYPPVVMHAHTVAVFLSFD
GDLPTQLEQXWRKAGKVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPV
DVLPSRLDVLTPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPV
LPVEWDPLPRAVTPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLP
6267..6659
/feature="O-island #118; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
6396..6476
/feature="24250"
6396..6476
/feature="24250"
/feature="orf; Unknown function"
/feature="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG58039.1"
/db_xref="GI:12517442"
/translation="MKWLSKSVQGDAPRIALIDIVDSL"
/complement(6814..8046)
/feature="serA"
/feature="24251"
/complement(6814..8046)
/feature="serA"
/feature="enzyme; Amino acid biosynthesis: Serine"
/feature="Residues 1 to 410 are 100.00 pct identical
to residues 1 to 410 of 410 from Escherichia coli K-12
Strain MG1655: B2913"
/codon_start=1
/transl_table=11
/product="D-3-phosphoglycerate dehydrogenase"
/protein_id="AAG58040.1"
/db_xref="GI:12517443"
/translation="MAKVSLEKDKIKFLVEGVHOKALESRLAAGTYNIEPHKGLDD
EOLKESIRDAHEIGLSRTHLTEDVINAELKIVGFCFICGTNOVDLDAARKGIPVF
NAPENSTSVAEVLIGELLILLRGVPEANAKAHGVWKNLAAGFEARGKGLIGIYG
HIGTOLGIALISLGMVYFYDIENKPLGNATQVORHSLDLNMSDVSLHVPENPSTK
Query Match          9.5%; Score 82.6; DB 1; Length 10367;
Best Local Similarity 48.8%; Pred. No. 1.7e-12;
Matches 223; Conservative 0; Mismatches 234; Indels 0; Gaps 0;
QY 35 tcattgataagcagcttcgaaggcgcctcttagcccttccattccctccctcgagg 94
Db 9386 TGATACGTGAACGAGGATTTGACGCGCGGACAAAGATGTGCAATACCAATCAGCGG 9445
QY 95 tgagtcagcgcgttaaagctctcgagcatcagctggtgagtggttgatcgcgaccc 154
Db 9446 TCTCAGACGCATTAGCAACTGGAATAATATTTCGGGACGCCGTGTTGGTACCG 9505
QY 155 aaccggccaaagcaaccgagcgggtgaagtccttggtgcaagcagcgcggaataatggtg 214
Db 9506 TACCGCGCGCGCGGACGGAACAGGCGCAAAACTGTGTCGACGTGCTGCGCAGGTGAGT 9565
QY 215 tctgcaagcaaaactaaagcgaactatctggagcctctggaactcgtgaaatcctgtaacca 274
Db 9566 TCTGGAAGAGAGTGGCTGGCGGATGGAACAAACCGTTTCGCTGCTGCTTTTCAC 9625
QY 275 tcgccatcaacgcagattcgctatccacatggtttctccctccctgttcaacgaggtagctt 334
Db 9626 TGGCGGTCAATCCGACAGCTTGGCGACGCTGTTGCTTCCGGCAGCTGGCTCTGATTGG 9685
QY 335 cttg999gtggaacagcctcagctcgcgtcgttgaagatgaagcgcacacattatccttgc 394
Db 9686 CTGATTGCGCTATCCGCCCAACTTTCAGGTAGAGATGAACCCGACACTCAGGAACGTC 9745
QY 395 tcg999gtgagatgttttagagcggtaaccgcggaactaatccggtg999gagatgtg 454
Db 9746 TCGTCCGCGGGAAGTGGTGGCGCGGTGAGTATTCACATCAGGCGCTCCGAGGTGTC 9805
QY 455 aagtagaagacttggaacccatgcgcacttggccat 491
Db 9806 TTGTCGATAAACTTGTGGCGCTCGACTATCTGTCGT 9842

```

RESULT 14  
AC020850\_0/c  
WPCOMMENT

Sequence split into 5 fragments LOCUS AC020850 Accession AC020850

Fragment Name	Begin	End
AC020850_0	1	110000
AC020850_1	100001	210000
AC020850_2	200001	310000
AC020850_3	300001	410000
AC020850_4	400001	444882

LOCUS AC020850 444882 bp DNA HTG 12-JUL-2000  
DEFINITION Mus musculus clone Rp21-460J8, WORKING DRAFT SEQUENCE, 179  
unordered pieces.

ACCESSION AC020850  
VERSION AC020850.3 GI:8810264  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 444882)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Mouse  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 444882)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Jun 29, 2000 this sequence version replaced gi:6980223.

-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 1434500  
Center clone name: RPCI-21\_460J8  
-----  
Summary Statistics  
Consensus quality: 319973 bases at least Q40  
Consensus quality: 373739 bases at least Q30  
Consensus quality: 387200 bases at least Q20  
Estimated insert size: 180000; pulse field gel estimation  
Estimated insert size: 427082; sum-of-contigs estimation  
Quality coverage: 10.25 in Q20 bases; pulse field gel estimation  
Quality coverage: 4.32 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 179 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1	1157: contig of 1157 bp in length
*	1158: gap of unknown length
*	1258: contig of 1148 bp in length
*	2405: gap of unknown length
*	2406: gap of unknown length
*	2506: contig of 1158 bp in length
*	3664: gap of unknown length
*	3763: contig of 1257 bp in length
*	5021: gap of unknown length
*	5121: contig of 1315 bp in length
*	6436: gap of unknown length
*	7636: contig of 1101 bp in length
*	7637: gap of unknown length
*	7737: contig of 1051 bp in length
*	8788: gap of unknown length
*	10443: contig of 1556 bp in length
*	10543: gap of unknown length
*	10544: contig of 1119 bp in length

```
* 11663 11762: gap of unknown length
* 11763 13127: contig of 1365 bp in length
* 13128 13127: gap of unknown length
* 13228 14743: contig of 1516 bp in length
* 14744 14843: gap of unknown length
* 14844 16088: contig of 1245 bp in length
* 16089 16188: gap of unknown length
* 16189 17328: contig of 1140 bp in length
* 17329 17428: gap of unknown length
* 17429 18672: contig of 1244 bp in length
* 18673 18772: gap of unknown length
* 18773 19972: contig of 1200 bp in length
* 19973 20072: gap of unknown length
* 20073 21352: contig of 1280 bp in length
* 21353 21452: gap of unknown length
* 21453 22678: contig of 1226 bp in length
* 22679 22778: gap of unknown length
* 22779 23972: contig of 1194 bp in length
* 23973 24072: gap of unknown length
* 24073 25586: contig of 1514 bp in length
* 25587 25686: gap of unknown length
* 25687 26748: contig of 1062 bp in length
* 26749 26848: gap of unknown length
* 26849 27933: contig of 1085 bp in length
* 27934 28033: gap of unknown length
* 28034 29541: contig of 1508 bp in length
* 29542 29641: gap of unknown length
* 29642 30965: contig of 1324 bp in length
* 30966 31065: gap of unknown length
* 31066 32222: contig of 1157 bp in length
* 32223 32322: gap of unknown length
* 32323 33365: contig of 1043 bp in length
* 33366 34494: contig of 1029 bp in length
* 34495 34594: gap of unknown length
* 34595 35652: contig of 1058 bp in length
* 35653 35752: gap of unknown length
* 35753 37270: contig of 1518 bp in length
* 37271 38377: gap of unknown length
* 38378 38477: gap of unknown length
* 38478 39510: contig of 1033 bp in length
* 39511 39610: gap of unknown length
* 39611 40635: contig of 1025 bp in length
* 40636 40735: gap of unknown length
* 40736 42202: contig of 1467 bp in length
* 42203 42302: gap of unknown length
* 42303 43468: contig of 1166 bp in length
* 43469 43568: gap of unknown length
* 43569 44658: contig of 1090 bp in length
* 44659 44758: gap of unknown length
* 44759 46187: contig of 1429 bp in length
* 46188 46287: gap of unknown length
* 46288 47312: contig of 1025 bp in length
* 47313 47412: gap of unknown length
* 47413 48737: contig of 1325 bp in length
* 48738 48837: gap of unknown length
* 48838 49905: contig of 1068 bp in length
* 49906 50006: gap of unknown length
* 50007 51130: contig of 1125 bp in length
* 51131 51230: gap of unknown length
* 51231 52484: contig of 1254 bp in length
* 52485 52584: gap of unknown length
* 52585 53823: contig of 1239 bp in length
* 53824 53923: gap of unknown length
* 53924 55321: contig of 1398 bp in length
* 55322 55421: gap of unknown length
* 55422 56688: contig of 1267 bp in length
* 56689 56788: gap of unknown length
* 56789 58069: contig of 1281 bp in length
* 58070 58169: gap of unknown length
* 58170 59706: contig of 1537 bp in length
* 59707 59806: gap of unknown length
```

```
* 59807 60981: contig of 1175 bp in length
* 60982 61081: gap of unknown length
* 61082 62213: contig of 1132 bp in length
* 62214 62313: gap of unknown length
* 62314 63704: contig of 1391 bp in length
* 63705 63804: gap of unknown length
* 63805 65223: contig of 1719 bp in length
* 65224 65623: gap of unknown length
* 65624 66821: contig of 1198 bp in length
* 66822 67957: contig of 1036 bp in length
* 67958 68057: gap of unknown length
* 68058 69361: contig of 1304 bp in length
* 69362 69461: gap of unknown length
* 69462 70606: contig of 1145 bp in length
* 70607 72136: contig of 1430 bp in length
* 72137 72236: gap of unknown length
* 72237 73864: contig of 1628 bp in length
* 73865 73964: gap of unknown length
* 73965 75021: contig of 1057 bp in length
* 75022 75121: gap of unknown length
* 75122 76427: contig of 1306 bp in length
* 76428 76527: gap of unknown length
* 76528 78149: contig of 1622 bp in length
* 78150 78249: gap of unknown length
* 78250 79627: contig of 1378 bp in length
* 79628 79727: gap of unknown length
* 79728 80797: contig of 1070 bp in length
* 80798 80897: gap of unknown length
* 80898 81926: contig of 1029 bp in length
* 81927 82026: gap of unknown length
* 82027 83037: contig of 1011 bp in length
* 83038 83137: gap of unknown length
* 83138 84501: contig of 1364 bp in length
* 84502 84601: gap of unknown length
* 84602 86130: contig of 1529 bp in length
* 86131 86230: gap of unknown length
* 86231 87400: contig of 1170 bp in length
* 87401 87500: gap of unknown length
* 87501 89067: contig of 1567 bp in length
* 89068 89167: gap of unknown length
* 89168 90312: contig of 1145 bp in length
* 90313 90412: gap of unknown length
* 90413 91856: contig of 1444 bp in length
* 91857 91956: gap of unknown length
* 91957 93077: contig of 1121 bp in length
* 93078 93177: gap of unknown length
* 93178 94294: contig of 1117 bp in length
* 94295 94394: gap of unknown length
* 94395 95952: contig of 1558 bp in length
* 95953 96052: gap of unknown length
* 96053 97383: contig of 1331 bp in length
* 97384 97483: gap of unknown length
* 97484 98598: contig of 1115 bp in length
* 98599 99832: contig of 1134 bp in length
* 99833 99832: contig of 1134 bp in length
```

Query Match 9.5%; Score 82.6; DB 66; Length 110000;

Best Local Similarity 48.8%; Pred. No. 1.5e-12;

Matches 223; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

Qy 35 tcattgatgaaggcagcttcgaagcgccctccttagcccttccattcccccgcgcg 94

Db 96909 TGATACGTGAACGAGGATTTGAGCGCGGCACAAAAGCTGTGCATTACACAATCAGCCG 96850

Qy 95 tgaagtcagcgcgttaagcctcgcagcatcacgtgggtcgagtggttgcgcgcaccc 154

Db 96849 TCTCACAGCGCATTTAGCAACTGGAAAAATATGTTCCGGGACGCCGCTGTTGGCTACCG 96790

Qy 155 aacgggcaaaagcaacgcagcggttaagtcctcttgcaagcagcgcaaaatgtgt 214

Db 96789 TACGGCGCGCCGACGGAACAGGGCAAAACTGCTGGGACTGCTGCGCCAGGTGGAGT 96730

[illegible]

/translation="MQSDVVAIVGGWVGLAVACGLOGSLRVAVLQRVQVPEPLAANA  
PQURVSAINAASEKYLRLGLGWODIILSRASCYHGMVMDKDSFGHISFDDQSMGY5  
HLGHVSIWIHYALMKAHQSSDITLLAPAELOQVAMGENETELTKDGSMLTARLV  
IGDAGANSWLRNKADIDITLFWYQHALVATIRTEPHDAVARQVHGEGLTFLPL5  
DPHLSIWSLPEEAQMOQASDEFENALNIAFDNRKLGLCKVESARQVFPPLTRYA  
ROFASRLALVGDAHTIHLAGOGVNLGFMDAAELIAELKRLHROCKDIGOYILRR  
YERSKRHSALMLAGMGFRDLFSGTNPAPKLLRDIGLKLADTLPGVKPQLIROMGL  
NDLPFWLR"  
complement(6731. .7909)  
/gene="ubiH"  
complement(6731. .7909)  
/gene="ubiH"  
/note="ORF\_f392"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA69075.1"  
/db\_xref="GI:882437"  
/translation="MSVILVGGWAGATLALATLSLHSGALPVHLIEATPESHAPG  
FDGRAIALAGTCQOLARIGVWQSLACATATITVHVSDRGHAGFVTLAAEDYQLAAL  
GQVVELHNSQRLUFALLRKAPGVTLHCPDRVANVARQSHVEVTESETLTGRVLVA  
ADGTHSALATACGYDMQOEPEYEQLAIVANVATSVAHGEGRAFERFQHGPLAMLPSDG  
RCSLVWCHPLERREVELSWDEKFCRELQSAFGWRLGKIITHAGKRKSAYPLALTHAARS  
ITHRTVLVGNAAOTLHPHAGOGFNLGMRDVMSLAETLTOAQERGEDMDGYGLCRYQO  
RROSDREATIGVTDLSLHVFANRWAPLVVGRNIGLTMELTTPARDVLAQRTLGWVAR  
"  
complement(7749. .9599)  
/note="similar to GenBank Accession Number D00398  
(ECOAPP2)"  
complement(7906. .9231)  
/gene="pepP"  
complement(7906. .9231)  
/gene="pepP"  
/note="ORF\_f441; third start codon"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA69076.1"  
/db\_xref="GI:882438"  
/translation="MSELSRQETQRRRQALVEQMOPGSAALIFAAPVTSADSEYFY  
RQNSDFWYTFNEPEALVLILKSDHNSVLFRNVRDLTAEIFRFRRLQGDAAPEK  
LGYDRALAFSEIQOLLYOLLVDVYHAGEXAYADVIIVNSALEKLRKSGRQNLTPAP  
ATMDWRPVVHEMLFKSPFEIIVLRRAGETIAMAHTFRAMEKCRPGFEXHLEGEIHH  
EFNRHGARYPSYNTIVGSGENGCIILHVTENECMRQGDVLIDAGCEYKYAGDITRT  
FVNGKFTQAKRELYDVLVLESLETSRLYRPGTSILEYGEVVRIMVSLVKLGLKG  
DVDELIAQNAHRFFHGLSHWGLDHDVYGDQDRSLRILEPGMVLVTEPGFLTIAPD  
AEVPEQYRGIGIEDIVITETGNENLTASVVKRPEEIEALMVAARKQ"  
complement(9257. .9841)  
/gene="ygfB"  
complement(9257. .9841)  
/gene="ygfB"  
/note="ORF\_f194"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA69077.1"  
/db\_xref="GI:882439"  
/translation="MLMSIONEMPGYNEMNOYLNQOGTGLTPAEMHGLISGMICGGND  
DSSWPLLLHDLTNEGHAFHGLAQALRKMHSATSDALQDGFLLYLPQDDVSVFDP  
RADLAGVWNHFLLGIGVTPQPKDKVGTGETGEAIDDLURNIAQLGYDEDEDELEMSL  
EETIEYVRVAALLCHDTFTHPQTAPEVQKPTLR"  
10003. .10332  
/note="ORF\_o109"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA69078.1"  
/db\_xref="GI:882440"  
/translation="MSAQPVDIQIFGRSLRVNCPDQDQALNQAAADLNORLQDLKER  
TRVNTQQLVFIAALNISYELAQEKAKTRDYAASMEQRIQLQOTIEQALLEQGRITE  
KTQNEP"  
10266. .11219  
/note="similar to GenBank Accession Number M12965  
(ECOSSR)"  
10374. .10556  
/gene="ssr"  
/product="stable 6S RNA"

gene 10374. .10556  
/gene="ssr"  
10374. .10557  
misc\_feature /note="similar to GenBank Accession Number X01238  
(ECRRNGS)"  
Query Match 9.5%; Score 82.6; DB 2; Length 141744;  
Best Local Similarity 48.8%; Pred. No. 1.4e-12;  
Matches 223; Conservative 0; Mismatches 234; Indels 0; Gaps 0;  
Qy 35 tcattgtaaggcagcttcgaagcgccctcttagcccttccattcccccctcgcgcg 94  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14187 TGATACGTGAACGAGGATTGAGCGCGCGGCACAAAACCTGTGCATTACAAATCAGCCG 14246  
Qy 95 tgagtcagcgcttaagctctcgagcatcacgtgggtcgaggtgtgtgtatcgcgcaccc 154  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14247 TCTCACGCGCATTAAGCAACTGGAAATAATGTTTCGGGCAGCGCGTGTGGTGCGTACCG 14306  
Qy 155 aaccggccaaagcaaccgaagcggtgaagtccttctgtgcaagcagcgcggaataatggtgt 214  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14307 TACCGCCGCGCCGACGGAACAAGGGCAAAAACCTGCTGGCACGTGCTGCCGCGCAGGTGAGT 14366  
Qy 215 tgctgcaagcagaaaaactaaagcgcaactatctggagcgtcttgctgaaatcccgtaaacca 274  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14367 TGCTGGAAGAAGAGTGGCTGGCGGATGAACAAACCGGTTCCACTCCGCTGCTTTTCAC 14426  
Qy 275 tcgcatcaacgcagattcgtctatccacatggtttctctccctcggttcaacaggtagctt 334  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14427 TGGCGGTCAACGCCGCACAGTCTGGCGAGCTGTTGCTTCTTCGACTGGCTCTGTTGTGG 14486  
Qy 335 cttgggggtggagcaacgctcagctgcgttggagatgaagatgaagcacaattatcccttgc 394  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14487 CTGATTTCGCCTATCCGCGCTCAACTTGCAGGTAGAGATGAACCCGCACTCAGGAACGTC 14546  
Qy 395 tgcggcgtggagatgttttagagcggttaaccccggtgaagctaatcccggtggcggtatgtg 454  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14547 TGCGCCGCGGGAAGTGGTCCGCGCGGTGAGTATTCAACATCAGGCGGCTGCCGAGTTGTC 14606  
Qy 455 aagtagtagaacttggaaacattgcccacttggccat 491  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14607 TTGTCGATAAACTTGGTCCGCTCGACTATCTGTTTCGT 14643  
Search completed: September 14, 2001, 07:55:07  
Job time: 15209 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2001, 07:54:24 ; Search time 171.99 Seconds  
(without alignments)  
3187.151 Million cell updates/sec

Title: us-09-105-117i-1-copy\_1421\_2293

Perfect score: 873

Sequence: 1 atgacccattcaactgga.....tcgagggttcgagccttag 873

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601:\*

- 1: /cgnl\_9/gcgdata/geneseq/geneseq/NA1980.DAT:\*
- 2: /cgnl\_9/gcgdata/geneseq/geneseq/NA1981.DAT:\*
- 3: /cgnl\_9/gcgdata/geneseq/geneseq/NA1982.DAT:\*
- 4: /cgnl\_9/gcgdata/geneseq/geneseq/NA1983.DAT:\*
- 5: /cgnl\_9/gcgdata/geneseq/geneseq/NA1984.DAT:\*
- 6: /cgnl\_9/gcgdata/geneseq/geneseq/NA1985.DAT:\*
- 7: /cgnl\_9/gcgdata/geneseq/geneseq/NA1986.DAT:\*
- 8: /cgnl\_9/gcgdata/geneseq/geneseq/NA1987.DAT:\*
- 9: /cgnl\_9/gcgdata/geneseq/geneseq/NA1988.DAT:\*
- 10: /cgnl\_9/gcgdata/geneseq/geneseq/NA1989.DAT:\*
- 11: /cgnl\_9/gcgdata/geneseq/geneseq/NA1990.DAT:\*
- 12: /cgnl\_9/gcgdata/geneseq/geneseq/NA1991.DAT:\*
- 13: /cgnl\_9/gcgdata/geneseq/geneseq/NA1992.DAT:\*
- 14: /cgnl\_9/gcgdata/geneseq/geneseq/NA1993.DAT:\*
- 15: /cgnl\_9/gcgdata/geneseq/geneseq/NA1994.DAT:\*
- 16: /cgnl\_9/gcgdata/geneseq/geneseq/NA1995.DAT:\*
- 17: /cgnl\_9/gcgdata/geneseq/geneseq/NA1996.DAT:\*
- 18: /cgnl\_9/gcgdata/geneseq/geneseq/NA1997.DAT:\*
- 19: /cgnl\_9/gcgdata/geneseq/geneseq/NA1998.DAT:\*
- 20: /cgnl\_9/gcgdata/geneseq/geneseq/NA1999.DAT:\*
- 21: /cgnl\_9/gcgdata/geneseq/geneseq/NA2000.DAT:\*
- 22: /cgnl\_9/gcgdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	873	100.0	993	22 AAF71777	Corynebacterium gl
2	873	100.0	2374	18 AAT96816	DNA encoding lysG,
3	122.6	14.0	15239	17 AAT33536	BCG deletion regio
4	45.4	5.2	45613	22 AAF28535	Genomic fragment #
5	45.2	5.2	4915	20 AAX84317	Stealth virus nucl
6	41.6	4.8	2842	21 AAD00334	Rice raffinose syn
7	37	4.2	2058	9 AAN80093	Sequence of BamHI/
8	37	4.2	2058	16 AAQ97857	P. aeruginosa tlda
9	36.8	4.2	102634	21 AAB81464	N. meningitidis pa
10	36.8	4.2	349980	21 AAF21609	Neisseria meningit
11	35	4.0	4689	21 AAT87299	S. venezuelae macr

C 12	35	4.0	36778	21	AAZ87318	S. venezuelae pik
C 13	35	4.0	37948	21	AAZ87285	S. venezuelae pik
C 14	35	4.0	38506	21	AAZ75633	Nucleotide sequenc
C 15	35	4.0	38506	21	AAZ56001	Recombinant cosmid
C 16	34.8	4.0	1431	8	AAZ70264	Sequence encoding
C 17	34.6	4.0	777	22	AAZ72046	Corynebacterium gl
C 18	34.6	4.0	777	22	AAZ72047	Corynebacterium gl
C 19	34.6	4.0	1704	20	AAZ99698	Human TGF-beta act
C 20	34.6	4.0	2785	20	AAZ99696	Human TGF-beta act
C 21	34.6	4.0	2866	20	AAZ99697	Human TGF-beta act
C 22	34.4	3.9	5819	20	AAZ20533	Polynucleotide seq
C 23	33	3.8	379	16	AAZ01036	Human B7-1 exon 5.
C 24	33	3.8	1491	15	AAZ05804	B7 antigen gene.
C 25	33	3.8	1491	16	AAZ081371	Human B lymphocyte
C 26	33	3.8	1491	20	AAZ000757	Human B7 cDNA sequ
C 27	33	3.8	1491	20	AAZ55786	Human B7-2 antigen
C 28	33	3.8	1491	21	AAZ84051	Human B lymphocyte
C 29	33	3.8	1491	21	AAZ61328	Human B7 cDNA. HO
C 30	33	3.8	1491	21	AAZ29320	Human B7.1 cDNA.
C 31	33	3.8	1491	21	AAZ40022	Human B7.1 coding
C 32	33	3.8	1491	22	AAZ89224	Human B lymphocyte
C 33	33	3.8	1788	20	AAZ56285	Human TAK1-6xHis e
C 34	33	3.8	1959	18	AAZ85095	Human transforming
C 35	33	3.8	2656	20	AAZ56279	Human TAK1 encodin
C 36	33	3.8	2656	21	AAZ39105	Human TAK-1 nucleo
C 37	33	3.8	7419	21	AAZ49922	Human calcium chan
C 38	32.4	3.7	1329	19	AAZ38205	Human mucosal adre
C 39	32.4	3.7	1536	19	AAZ38201	Human mucosal adre
C 40	32.4	3.7	1539	17	AAZ39751	Human mucosal addr
C 41	32.4	3.7	1539	18	AAZ35280	cDNA encoding huma
C 42	32.4	3.7	1546	21	AAZ21163	Human low adenosin
C 43	32.4	3.7	1546	21	AAZ35041	Human adenosine re
C 44	32.4	3.7	2868	21	AAZ00332	Corn raffinose syn
C 45	32.4	3.7	5539	21	AAZ21164	Human low adenosin

#### ALIGNMENTS

#### RESULT 1

AAZ71777  
ID AAF71777 standard; DNA; 993 BP.

XX AAF71777;

XX 30-APR-2001 (first entry)

XX Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:49.

XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
fine chemical production; microorganism; organic acid; nucleoside;  
nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.

XX Corynebacterium glutamicum.

XX WO200100843-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB00923.

XX 25-JUN-1999; 99US-0141031.

XX 01-JUL-1999; 99DE-1030476.

XX 02-JUL-1999; 99US-0142101.

XX 08-JUL-1999; 99DE-1031415.

XX 08-JUL-1999; 99DE-1031418.

XX 08-JUL-1999; 99DE-1031419.

XX 08-JUL-1999; 99DE-1031420.

XX 08-JUL-1999; 99DE-1031424.

XX 08-JUL-1999; 99DE-1031428.

XX 08-JUL-1999; 99DE-1031434.

PR 08-JUL-1999; 99DE-1031435.  
 PR 08-JUL-1999; 99DE-1031443.  
 PR 08-JUL-1999; 99DE-1031453.  
 PR 08-JUL-1999; 99DE-1031457.  
 PR 08-JUL-1999; 99DE-1031465.  
 PR 08-JUL-1999; 99DE-1031478.  
 PR 08-JUL-1999; 99DE-1031510.  
 PR 08-JUL-1999; 99DE-1031541.  
 PR 08-JUL-1999; 99DE-1031573.  
 PR 08-JUL-1999; 99DE-1031592.  
 PR 08-JUL-1999; 99DE-1031632.  
 PR 08-JUL-1999; 99DE-1031634.  
 PR 08-JUL-1999; 99DE-1031636.  
 PR 09-JUL-1999; 99DE-1032125.  
 PR 09-JUL-1999; 99DE-1032126.  
 PR 09-JUL-1999; 99DE-1032130.  
 PR 09-JUL-1999; 99DE-1032186.  
 PR 09-JUL-1999; 99DE-1032206.  
 PR 09-JUL-1999; 99DE-1032227.  
 PR 09-JUL-1999; 99DE-1032228.  
 PR 09-JUL-1999; 99DE-1032229.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 14-JUL-1999; 99DE-1032922.  
 PR 14-JUL-1999; 99DE-1032926.  
 PR 14-JUL-1999; 99DE-1032928.  
 PR 14-JUL-1999; 99DE-1033004.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 12-AUG-1999; 99US-0148613.  
 PR 27-AUG-1999; 99DE-1040764.  
 PR 27-AUG-1999; 99DE-1040765.  
 PR 27-AUG-1999; 99DE-1040766.  
 PR 27-AUG-1999; 99DE-1040832.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 31-AUG-1999; 99DE-1041380.  
 PR 31-AUG-1999; 99DE-1041394.  
 PR 31-AUG-1999; 99DE-1041396.  
 PR 03-SEP-1999; 99DE-1042076.  
 PR 03-SEP-1999; 99DE-1042077.  
 PR 03-SEP-1999; 99DE-1042079.  
 PR 03-SEP-1999; 99DE-1042086.  
 PR 03-SEP-1999; 99DE-1042087.  
 PR 03-SEP-1999; 99DE-1042088.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042124.  
 PR 03-SEP-1999; 99DE-1042129.  
 PR 09-MAR-2000; 2000US-0187970.  
 PA (BADI ) BASF AG.

XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
 PI WPI: 2001-137957/14.  
 DR P-PSDB; AAB79658.  
 XX  
 PT Nucleic acids from *Corynebacterium glutamicum* encoding metabolic  
 PT pathway proteins, useful for producing fine chemicals in  
 PT microorganisms, including organic acids, nonproteinogenic amino acids,  
 PT and purine and pyrimidine bases -  
 XX  
 PS Claim 3: Page 226-228; 1737pp; English.  
 XX

CC AAF1753 to AAF72330 encode the *Corynebacterium glutamicum* metabolic  
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
 CC MP nucleic acids are useful for the production of fine chemicals  
 CC in microorganisms, including organic acids, nonproteinogenic amino  
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,  
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic  
 CC compounds, vitamins, cofactors, polyketides and enzymes.  
 XX  
 SQ Sequence 993 BP; 233 A; 256 C; 268 G; 236 T; 0 other;

Query Match		100.0%;	Score 873;	DB 22;	Length 993;
Best Local Similarity		100.0%;	Pred. No. 1.9e-268;		
Matches	873;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	1	atgaacccattcaactgacactttgctctcaatcattgatgaagcgcttcgaagc	60		
DB	101	atgaacccattcaactgacactttgctctcaatcattgatgaagcgcttcgaagc	160		
QY	61	gcctcttagcctttccatttccctcctcgcggtgagtcagcgctttaaagctctcgag	120		
DB	161	gcctcttagcctttccatttccctcctcgcggtgagtcagcgctttaaagctctcgag	220		
QY	121	catacgtgggtcgagtgttggtatcgcgcaacccaccgccaacacccggaagcggt	180		
DB	221	catacgtgggtcgagtgttggtatcgcgcaacccaccgccaacacccggaagcggt	280		
QY	181	gaagtccttgcgaagcagcgcggaataatgggttctcgaagcagaactaaagcgaa	240		
DB	281	gaagtccttgcgaagcagcgcggaataatgggttctcgaagcagaactaaagcgaa	340		
QY	241	ctatctggagccttgcgaagcagcgcggaataatgggttctcgaagcagaactaaagcgaa	300		
DB	341	ctatctggagccttgcgaagcagcgcggaataatgggttctcgaagcagaactaaagcgaa	400		
QY	301	acatggtttctcctccgtgttcaacgaggtagcttcttgggtgagcaacgctcacgctg	360		
DB	401	acatggtttctcctccgtgttcaacgaggtagcttcttgggtgagcaacgctcacgctg	460		
QY	361	cgcttgaagatgaagcagcagcattatcttctgctcgcggtggaagatgttttagagcg	420		
DB	461	cgcttgaagatgaagcagcagcattatcttctgctcgcggtggaagatgttttagagcg	520		
QY	421	gtaacccgtgaagcagcagcattatcttctgctcgcggtggaagatgttttagagcg	480		
DB	521	gtaacccgtgaagcagcagcattatcttctgctcgcggtggaagatgttttagagcg	580		
QY	481	cacttgccattgcaacccctccatcttgcgggagtgctacatggttgatgggaactagat	540		
DB	581	cacttgccattgcaacccctccatcttgcgggagtgctacatggttgatgggaactagat	640		
QY	541	tgggtgcgactccctctttagccttcggtcccaagatgtcttcaagacgctgacctg	600		
DB	641	tgggtgcgactccctctttagccttcggtcccaagatgtcttcaagacgctgacctg	700		
QY	601	gacggcgcgctcgatgctctgtggcgcgagcgatccatttcccgctcgcggaag	660		
DB	701	gacggcgcgctcgatgctctgtggcgcgagcgatccatttcccgctcgcggaag	760		
QY	661	ggtttgtgaggaattcgccgagccttgggtgggagcttcttcccgaaacccaagct	720		
DB	761	ggtttgtgaggaattcgccgagccttgggtgggagcttcttcccgaaacccaagct	820		
QY	721	gctccatgctaaagcaggaagtgatccctctcgatgagatccattacacacg	780		
DB	821	gctccatgctaaagcaggaagtgatccctctcgatgagatccattacacacg	880		
QY	781	atgtattgcaacgattggcgcttgaatcttagctcttagctacacacgctc	840		
DB	881	atgtattgcaacgattggcgcttgaatcttagctcttagctacacacgctc	940		
QY	841	gttgatgcagcaatcgaggattggcgcttag	873		
DB	941	gttgatgcagcaatcgaggattggcgcttag	973		
RESULT 2					
AAT96816/C					
ID AAT96816 standard; DNA; 2374 BP.					
XX					
AC AAT96816;					
XX					
DT 12-MAR-1998 (first entry)					

XX DNA encoding LysG, LysE and ORF3 from Corynebacterium glutamicum.  
 DE  
 XX  
 KW LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;  
 KW Microbial production; amino acid; animal feed additive; ds.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS complement (82..954)  
 FT /\*tag= a  
 FT /label= LysG  
 FT 1016..1726  
 FT /\*tag= b  
 FT /label= LysE  
 FT complement (1723..2373)  
 FT /\*tag= c  
 FT /label= orf3  
 XX  
 PN DEL9548222-AL.  
 XX  
 PD 26-JUN-1997.  
 XX  
 XX 22-DEC-1995; 95DE-1048222.  
 XX  
 XX 22-DEC-1995; 95DE-1048222.  
 PR  
 XX (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.  
 PA  
 XX Eggeling L, Sahm H, Vrljic M;  
 PI  
 XX WFI; 1997-333867/31.  
 DR  
 DR P-PSDB; AAW37714-16.  
 XX  
 XX Increasing microbial production of amino acids, especially lysine -  
 PT by improving export carrier activity or corresponding gene  
 PT expression, also new export and regulatory genes from  
 PT Corynebacterium  
 XX  
 XX Claim 23 and 26; Page -: 16pp; German.  
 PS  
 XX This DNA, isolated from Corynebacterium glutamicum, contains the LysG,  
 CC LysE and ORF3 genes. LysG and LysE encode a lysine transport regulatory  
 CC protein and an export protein, respectively. Microbial production of  
 CC amino acids (A) is improved by increasing the export-carrier activity  
 CC and/or the export gene expression in a microorganism that produces (A).  
 CC The method is specifically used to increase production of lysine,  
 CC used as an animal feed additive. Other (A) are variously useful as  
 CC pharmaceuticals, condiments and intermediates for fine chemicals.  
 CC This method increases the amount of (A) secreted into the culture medium.  
 CC Export of (A) has been found to depend on a single gene.  
 CC NB. This sequence has been created from the information given in table 2  
 CC of the specification.  
 XX  
 XX Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other;  
 SQ

Query Match 100.0%; Score 873; DB 18; Length 2374;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-268;  
 Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgaacccattcaactggacactttgtctcattatcattgatgaagcagcttcgaaggc 60  
 |||||  
 Db 954 ATGAACCCATTCACTGGACATTTGCTCTCAATCATTTGATGAAGCAGCTTCGAAGGC 895  
 |||||  
 Qy 61 gctccttagccctttccattccctcctcgccggtgagtcagcgcgttaagctctcag 120  
 |||||  
 Db 894 GCCTCCTTAGCCCTTTCCATTTCCCTCGCGGTGAGTCAGCGCGTTAAAGCTCTCGAG 835  
 |||||  
 Qy 121 catcacgtggtcagtggttgatcgcacccacccgacccgacccgacccgacccggt 180  
 |||||  
 Db 834 CATCACGTGGTGTGAGTGTGTTATCGCGACCCACCCGCGCAACCGCAACCGAGCGGGT 775

Qy 181 gaatccttctgcaagcagcgcggaaaaatggtgtgtgtgcaagcagaactaaagcga 240  
 |||||  
 Db 774 GAATCCTTGTGCAAGCAGCGCGGAAAATGTTGTGTGCAAGCAGAACTAAAGCGCAA 715  
 |||||  
 Qy 241 ctatctgagccttgcgtgaatcccgtaaccatcccatccatcaacgcagatctcgtatcc 300  
 |||||  
 Db 714 CTATCTGGACGCTTGTGAAATCCCGTTACCATGCCATCAACGCAGATTCCTATCC 655  
 |||||  
 Qy 301 acatggtttcctccctggttcaacgaggtagcttcttgggtggagcaacgcctcacctg 360  
 |||||  
 Db 654 ACATGGTTTCTCTCCGTTCAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCACGCTG 595  
 |||||  
 Qy 361 cgttggagaatgaagcgcacacattatccttgcgtcggcggtggagatgttttagagcg 420  
 |||||  
 Db 594 CGCTTGAAGATGAAGCGCACACATTATCTCTGCTGGCGGTGGAGATGTTTATAGGAGCG 535  
 |||||  
 Qy 421 gtaaccgtgaagcctaataccgctggtggatgtgaagttagtagaacttggaaacctgcgc 480  
 |||||  
 Db 534 GTAACCGGTGAAGCTAATCCGTCGCGGATGTGAAGTAGTAGAAGCTTGGAAACCATGGC 475  
 |||||  
 Qy 481 cacttggccattgcaacccctcattcggggtgcctacatggttgcgtgggaaactagat 540  
 |||||  
 Db 474 CACTTGGCCATTGCAACCCCTCATTTGGGGATGCCTACATGTTGATGGGAAACTAGAT 415  
 |||||  
 Qy 541 tgggctcgatgccttcttacgcttcggtcccaaatgtcttcaagaccgtgacctg 600  
 |||||  
 Db 414 TGGGCTCGATGCGCGCTTACGCTTCGGTCCCAAGATGTGCTCAAGACCGTGACCTG 355  
 |||||  
 Qy 601 gacgggcgctcgatggtcctgtgggcgagcgcgctatccattgtccgctcgcgga 660  
 |||||  
 Db 354 GACGGCGCTCGATGCTGCTGTGGGGCGCAGGCGCGTATCCATTGTCCTCGCGCGGAA 295  
 |||||  
 Qy 661 ggtttggtgaggaattccgagcgttgggtgggtgggtggttgcgtgggaaactagct 720  
 |||||  
 Db 294 GGTTTTGTGAGGCAATTCCGCGAGGCTTGGTTGGGACTTCTTCCGAAACCAAGCT 235  
 |||||  
 Qy 721 gctccatgctaaaagcagagagaagtatcctcctcgtatgagataccattgacacaccg 780  
 |||||  
 Db 234 GCTCCCATGCTAAAGCAGGAGAGTAGTATCCTCTCGATGAGATACCCATTTGACACACCG 175  
 |||||  
 Qy 781 atgtattggaacatgagcgcctggaatctagatctctagtagactcacagacgcctc 840  
 |||||  
 Db 174 ATGATTGGCAACATGGCGCTGGAAATCTAGATCTCTTAGCTAGACTACAGACCGCGTC 115  
 |||||  
 Qy 841 gttgatcagcaatcaggggatttcgagccttag 873  
 |||||  
 Db 114 GTTCATGCAAGCAATCGAGGAGATTGCGGCTTAG 82

RESULT 3  
 AAT33536  
 ID AAT33536 standard; DNA; 15239 BP.  
 XX  
 AC -AAT33536;  
 XX  
 DT 15-FEB-1998 (first entry)  
 XX  
 DE BCG deletion region 2 and flanking sequences.  
 XX  
 KW BCG delta 2; virulence; avirulence; attenuation; gene deletion;  
 KW mycobacteria; vaccine; infection; marker; ss.  
 XX  
 OS Mycobacterium bovis strain BCG.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 3382..14071  
 FT /\*tag= a  
 FT /note= "BCG delta 1 deletion region"  
 XX  
 XX WO9625519-A1.  
 XX  
 XX 22-AUG-1996.  
 XX

```

PF 15-FEB-1996; 96WO-US01938.
XX
PR 17-FEB-1995; 95US-0390878.
XX
PA (PATH-) PATHOGENESIS CORP.
XX
PI Mahairas GG, Stover CK;
XX
DR WPI; 1996-393419/39.
XX
PT Detecting markers for avirulence in Mycobacterium - used in
PT production of vaccines against bacterial infection, and to detect
PT bacterial infection
XX
PS Example 1; Fig 2; 66pp; English.
XX
CC This DNA sequence comprises Mycobacterium bovis BCG deletion
CC sequence BCGdelta2. A specific genetic deletion of this region
CC results in an avirulence phenotype of the mycobacterium. 2 Other
CC deletion regions (see AAF28535 and AAF28537) have also been detected.
CC Identification involved screening a BCG cosmid library with a
CC radiolabeled probe obtained following DNA subtraction between
CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
CC The deletions provide useful markers for the identification of an
CC avirulent, or a virulent, mycobacterial phenotype. Determination
CC of avirulence requires the detection of the presence or absence of
CC the deletion; the deletions are detected either by detecting the
CC presence or absence of deletion junctions (see AAF28538-46), or by
CC detecting the presence or absence of the sequences contained within
CC the deletion. Deletion polypeptides are used as components of
CC immunological assays and in vaccines.
XX
SQ Sequence 15239 BP; 2878 A; 4791 C; 4729 G; 2841 T; 0 other;

Query Match 14.0%; Score 122.6; DB 17; Length 15239;
Best Local Similarity 49.9%; Pred. No. 1.7e-28;
Matches 426; Conservative 0; Mismatches 409; Indels 18; Gaps 4;

QY 13 caactggacacttctcctcaatcattgaaagcagcttgcgaagcgcctcctcttaggc 72
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5253 cagctggcgcattgctgcctggttcgaactggcagcttcgatgcgcccgcgagcgc 5312
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 73 ctttcattccctccctcggcgtgagtcagcgcgttaaaagctctgcagcatcagctgggt 132
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5313 ctacatgtcacccctcggctcagtcagcagcatcaatcagcttgagcagcagctcggc 5372
QY 133 cgaagtgtgattcgcgcacccaacccgccaagcaacccggaacgcgggtgaaagtctgtg 192
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5373 cagggtgctggtgaggaagccatgctcgggcagcagccgaggtatcccgctgttg 5432
QY 193 caagcagcgcgggaaatggtgttcgcaagcagaaactaaagcgcgaactatctgg--ac 250
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5433 cagttggcgcgcgaacacagctgttcgagtcgcgagcgcgtcgtgaaatgggtggcaac 5492
QY 251 gcttgctgtaa----atcccgtaaccatcgccatcaacgcagagctcgtatcccaatgg 306
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5493 ggcgttgtaaacgcacgcggtaccattcgggttaaacgcgattccatggcgacatgg 5552
QY 307 tttctcccgctgtcaacgaggtagctcttctgggtggagcaacgctcagcgtgcgttg 366
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5553 ttttcgcccgtgttcgacg-----gtctcggcagctcgtcgtcagcttcggatc 5603
QY 367 gaagtgaagcgcacacattatcttctgctgcggtggagatgttttagagcgtgaacc 426
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5604 gaggaccagaccattccgcgcgtcgtcagcggaggtgtggtgcgtggcgcgtgacc 5663
QY 427 cgtgaagcctaaccggtggcggatgtgaagttagtaaaccttggaaccatgcgcaccttg 486
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5664 accgagcgaaccccggtgcgggtgcgggtgcaccccgctgggtgaaatgcgtaccta 5723
QY 487 gcaattgaacccctcattcggcgatgctacatggttgatgggaaactagatggcgt 546
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

## RESULT 4

AAF28535/c

ID AAF28535 standard; DNA; 45613 BP.

XX

AC AAF28535;

XX

DT 04-APR-2001 (first entry)

XX

DE Genomic fragment #22.

XX

KW Genomic library; bacteria; human upper airway; otitis media; sinusitis; bronchopulmonary; endocarditis; meningitis; ss.

XX

OS Moraxella catarrhalis.

XX

PN WO200078968-A2.

XX

PD 28-DEC-2000.

XX

PF 16-JUN-2000; 2000WO-US16649.

XX

PR 18-JUN-1999; 99US-0140121.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Lagace RE, Patterson C, Berg KL;

XX

DR WPI; 2001-041427/05.

XX

PT Genomic library for identifying diagnostic and therapeutic

PT compositions, and for identifying virulence factors, regulatory

PT elements and drug targets, comprises Moraxella catarrhalis nucleic

PT acids.

XX

PS Claim 1; Page 180-191; 545pp; English.

XX

CC The present invention relates to a Moraxella catarrhalis genomic library comprising of a combination of 41 nucleic acid molecules (see

CC AAF28514-AAF28554). The library has a number of uses described in the

CC specification e.g. is useful for identifying diagnostic and therapeutic

CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large

CC aerobic, gram-negative diplococcus, normally found among the bacterial

CC flora of human upper airways. M. catarrhalis is known to cause acute,

CC localised infections such as otitis media, sinusitis and bronchopulmonary

CC infection and life-threatening, systemic diseases including endocarditis

CC and meningitis.



Query Match	4.88;	Score 41.6;	DB 21;	Length 2842;
Best Local Similarity	50.59;	Pred. No. 0.0054;		
Matches 101;	Conservative 0;	Mismatches 99;	Indels 0;	Gaps 0;
Qy	565	ttcgggtcccaagaatgtgtcttcaagaccgtgacctggacgcgcgctgcgatggctcctgtg	624	
Db	982	TTGCTCCGTGTCGTCTCCTCGGTGCCGATCTGTGCCAGCGCTCGATGATGAGGAA	923	
Qy	625	ggggcgagggcgctatccatttcccgtcgcggaagtgtttgtgagggcaattcgccga	684	
Db	922	CGCGGGCGCGCGCGCCATTGGCGAGGCTCGGAGGCGCTCTGACCGCGCTCGCGGT	863	
Qy	685	ggccttggttggggaacttctcccgaaacccaagctgtctcccatgctaaagcaggagaa	744	
Db	862	GACGTGCGTGTAGAACGGCGTCCCATGTGCACCCCGAACCAATCCAAAGAGGACGGCAT	803	
Qy	745	gtgatcctcctcctgatgagat	764	
Db	802	CTTCTTCTTCTTCCCTGTGTGT	783	

RESULT	7
AAN80093/c	
ID	AAN80093 standard; DNA; 2058 BP.
XX	
XX	
XX	AAN80093;
XX	
DT	30-MAR-1992 (first entry)
XX	
DE	Sequence of BamHI/Sali fragment of the tfdA gene which encodes 2,4-dichlorophenoxyacetic acid (2,4-D) mono-oxygenase.
DE	
DE	
XX	Herbicide resistance; halogenated aromatic cpd; enzyme;
KW	mono-oxygenase; ss.
KW	
XX	
OS	Alcaligenes eutrophus JMP134.
XX	
XX	
PH	Key Location/Qualifiers
FT	CDS 748..1608
FT	/*tag= a
FT	/note= "DNA SQ claimed"
XX	
XX	
PN	DE3629890-A.
XX	
PD	10-MAR-1988.
XX	
XX	
PF	29-AUG-1986; 86DE-3629890.
XX	
XX	
PR	29-AUG-1986; 86DE-3629890.
XX	
XX	(SCHD ) SCHERING AG.
PA	(STRE/) STREBER W R.
PA	(MACQ-) MACQUARIE UNIV.
XX	
PI	Streber W, Timmis KN, Zenk M;
XX	
XX	WPI; 1988-071716/11.
DR	P-PSDB; AAP80079.
DR	
XX	
PT	Plasmids and bacteria contg. gene tfdA for 2,4-D-mono:oxygenase -
PT	Isolated using new Alcaligenes eutrophus transposon mutants
XX	
PS	Disclosure; Fig 10; 32pp; German.
XX	
CC	The tfdA gene product is a 2,4-D decomposing protein. A.eutrophus
CC	JMP 134 contains an 80 kb plasmid, pJP4, which carries the genes
CC	coding for 2,4-D decomposition. Plasmids and bacteria contg. the
CC	gene are useful for prodn. of 2,4-D mono-oxygenase and for
CC	transferring the ability to decompose 2,4-D to other micro-organisms
CC	(which could then be useful for cleaning waste water polluted by
CC	halogenated aromatic cpds.) or to plants (to impart tolerance to 2,4-D).
XX	

SQ	Sequence 2058 BP; 372 A; 659 C; 661 G; 366 T; 0 other;
	Query Match            4.2%; Score 37; DB 9; Length 2058; Best Local Similarity 53.9%; pred. No. 0.14; Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
Qy	13 caactggacactttgctcctaattgatgaagcgagcttcgaagcgccctcttagacc 72 
Dd	530 CAGCTTCGCTATTTCGTGCGCCGGAGAGAGGCGACACTGGTGCCTCCGCCGCGCGG 471
Qy	73 ctttccatctccccctcgcgcgttgagtcaagcgcgttaaaagctctctcgagcatcaagtgggt 132 
Dd	470 CTGCATATTTCCAGCCCCCGGTACCGCAGAGATTTCACGGCGCTCGAACAGCATCTGGGC 411
Qy	133 cgagtgttgttgatcgcgcacc 153 
Dd	410 GTGTTGTGTTTCGAGCGCAGC 390

RESULT	8	
ID	AAQ97857/c	
AD	AAQ97857 standard; DNA; 2058 BP.	
XX	AC	
XX	AAQ97857;	
XX		
DT	06-DEC-1995 (first entry)	
XX		
DE	P. aeruginosa tfdA gene.	
XX		
XX	Transgenic plant; tfdA gene; 2,4-dichlorophenoxyacetic acid; 2,4-D;	
KW	herbicide resistance; sweetgum; Liquidambar styraciflua; hardwood;	
KW	plantation; crop improvement; selectable marker; pUCW200; vector;	
KW	ss.	
XX		
OS	Pseudomonas aeruginosa.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	751..1611
FT		/*tag= a
XX		
PN	W09518862-A.	
XX		
PD	13-JUL-1995.	
XX		
PF	10-JAN-1995; 95WO-US00284.	
XX		
PR	15-DEC-1994; 94US-0358117.	
PR	11-JAN-1994; 94US-0179667.	
XX		
PA	(UNBC ) UNION CAMP CORP.	
XX		
PI	Kaphammer BJ;	
XX		
XX	WPI: 1995-255064/33.	
DR	P-PSDB; AAR79659.	
XX		
PT	Using presence of tfdA gene to select transgenic plants - imparting	
PT	resistance to 2,4-di:chloro:phenoxy:acetic acid, esp. sweetgum	
PT	(Liquidambar) trees contg. this gene and plantations of them	
XX		
PS	Disclosure; Fig.7; 56pp; English.	
XX		
CC	A tfdA gene of P. aeruginosa plasmid pRO101, which encodes the	
CC	enzymes for 2,4-D degradation, was cloned into Agrobacterium	
CC	tumefaciens vector pBil121 to form pUCW200. This was used to	
CC	introduce 2,4-D herbicide-resistance into sweetgum, thereby	
CC	allowing selection of transgenic plants and reducing the cost of	
CC	sitting prepn. and maintenance.	
XX		
SQ	Sequence 2058 BP; 372 A; 659 C; 661 G; 366 T; 0 other;	

CC particularly organism-specific probes. Attempts to make efficacious  
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
CC Multivalent vaccines have also been tried but none have successfully  
CC overcome antigenic variability. The provision of further, complete  
CC sequences may provide an opportunity to identify secreted or surface  
CC exposed proteins that may be presumed targets for the immune system and  
CC which are not antigenically variable or at least more conserved than  
CC other more variable regions.

XX

SQ Sequence 102634 BP; 23871 A; 24828 C; 27888 G; 26042 T; 5 other;

Query Match 4.2%; Score 36.8; DB 21; Length 102634;  
Best Local Similarity 52.6%; Pred. No. 0.99;  
Matches 80; Conservative 0; Mismatches 72; Indels 0; Gaps

QY 1 atgaacccattcaactgagcacatttgctcctaattcataagaagcgacttcgaaggc 60  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 62462 atggatgcgcgtacaattaaaatacatcttgcgccgcgcgcagagggcaacctacccaa 62521  
QY 61 gctcctcttagcctttccatttccccctcggcggtgagtcagcgcgctaaagctctcgag 120  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 62522 gcgcgcaaacgactttctcttcccagcctgcgcttctgcccaaattaagcccttgaa 62581  
QY 121 catcagtgggtcgagtggtgatcgcgcac 152  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 62582 gaatatgcgcgcacgcgcgtgttcagcgcgac 62613

RESULT 10  
AAF21609  
ID ID AAF21609 standard; DNA; 349980 BP.  
XC AAF21609;  
XX  
DT 13-MAR-2001 (first entry)  
XX  
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:110.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
KN diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
KW ds.  
XX  
OS Neisseria meningitidis.  
PN WO200066791-A1.  
XX  
PD 09-NOV-2000.  
XX  
PF 08-MAR-2000; 2000WO-US05928.  
XX  
PR 30-APR-1999; 99US-0132068.  
PR 08-OCT-1999; 99WO-US23573.  
PR 28-FEB-2000; 2000GB-0004695.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;  
PI Galisetti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;  
PI Frazer CM, Grandi G;  
XX  
DR WPI; 2000-647603/62.  
XX  
XX Neisseria meningitidis B full length genome sequence and open reading  
PT frames are used to detect, treat and prevent Neisserial infections -  
PT  
PS Claim 7; Appendix A; 692pp; English.  
XX  
XX The present invention describes the full length genome of  
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607  
CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
CC sequence was too long to go in a record on its own it was split into 8

sequences which overlap each other at the beginning and end of each sequence by 4980 bp (i.e. the last 4980 bp of AAF21344 is repeated at the beginning of AAF21607, the last 4980 bp of AAF21547 is repeated at the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the *Neisseria* proteins given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR primers which are used in the exemplification of the present invention. The NMB genome and fragments from it have antibacterial activity, and can be used in vaccines and gene therapy. *Neisseria* nucleic acids, proteins and/or antibodies which binds to the proteins can be used in compositions for treating or preventing infection due to *Neisseria* bacteria or as a diagnostic reagent for detecting the presence of *Neisseria* bacteria or of antibodies raised to *Neisseria* bacteria. Computers, computer memory, computer storage medium or computer databases can be used in a search to identify open reading frames (ORFs) or coding sequences within the NMB genome. The DNA sequences provide further opportunities to find antigenic or immunogenic proteins which are more effective in vaccines than the outer membrane proteins currently used.

AA  
SQ Sequence 349980 BP; 81351 A; 86755 C; 95584 G; 86290 T; 0 other;

Query Match	4.28;	Score 36.8;	DB 21;	Length 349980;
Best Local Similarity	52.16%;	Pred. No. 1.8;		
Matches	80;	Conservative	0;	Mismatches 72;
				Indels 0;
				Gaps 0

Qy 1 atgaaccccatccaactggacactttgtctctcaatcattgatgaagcagcttcgaaggc 60  
 |||||  
 Db 166174 atggatgcgcgtacaataataataatattgttcgcgcgtcgcacacgagggcaaccttacccca 166233  
 |||||

[illegible]

**Qy**      121 catcacgagggtcgagtgcttggatcgcac 152  
          |         |         |         |         |  
**Pb** 166294 gaatatgcggcaaccgcgtgttcaggcgcac 166325

**RESULT 11**

AAZ8/299/C  
ID AAZ87299 standard; DNA; 4689 BP.

XX AC AAZ87299:

XX 05-JUN-2000 (first entry)

XX DE 2000

Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,  
chronic obstructive pulmonary disease; respiratory inflammation;  
hypercholesterolaemia; crop protection agent; ds.

OS *Streptomyces venezuelae* ATCC15439.

XX	Key	Location/Qualifiers
FH		

FT CDS 1..4689 /stage a

```
FT yy /product= "PikAIII"
```

PN WO20000620-A2.

PD 06~JAN-2000

PF 25-JUN-1999; 99WO-US14398.

PR 26-JUN-1998; 98US-0105537.

XX (MINU ) UNIV MINNESOTA. PA

XX  
4PT Sherman DU rin u yue v zhaC r.

DT 05-JUN-2000 (first entry)

XX DR DR XX PT PT XX PS XX CC

Query Match 4.0%; Score 35; DB 21; Length 4689;  
Best Local Similarity 45.7%; Pred. No. 0.87;

QY	406	gatgttttagagcgctaaccgcgtgaagctaaccgcgtggcggaatccgaagtagtagaa	465
Db	2991	GAGGTCGTCGAGGAGCGAGACACCGCCGGTGAAGCGTCGCGCGTCGTCAGTGCGGTGAG	2932

Qy 466 cttggaaccatgcgccacttgcccatgcacccccctcaattgcgggaatgcctacatggtt 525  
2931 CCGGGCGCGAGGGCCCTACGGTCTGCTGCTCCGCGCGCTTCCAGTACGTGCACCTCCGC 2872  
Db

Qy	585	gatgggaactagatgaggctgcgatacccggtcttaccgcttcacggtccccaaagtgtgctt
Db	2812	tcccgcggcctttacgccgcgccagcacgctcgctcttgctcgctccgacggcgacgat

Qy	586	caagaccgtgaactgacgaggcgctgcgatgtctctctgggcgcgagcgcgctatccatt	645
Db	2811	ccaccgcccgcagacgtccgctggccggaggccgctcagcggcttcacatcgacgcggta	2752

Qy 646 gtcccgctcgcggaaggttttggtgag 672  
 Db 2751 ACGCCAGTCGTCGGCGGGCGCTGGTGGG 2721

RESULT 12

AAZ8/318/C  
ID AAZ87318 standard; DNA: 36778 BP.

XX  
AC

XX  
DT 05-JUN-2000 (first entry)





XY

PF 27-MAY-1999; 99US-0320878.

XX

XX



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2001, 06:59:12 ; Search time 1873.17 Seconds  
(without alignments)  
4405.548 Million cell updates/sec

Title: US-09-105-117I-1-COPY\_1421\_2293  
Perfect score: 873  
Sequence: 1 atgaacccattcaactgga.....tcgagggttgccgcttag 873

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*

em\_esthum10:\*  
em\_esthum11:\*  
em\_esthum12:\*  
em\_esthum13:\*  
em\_esthum14:\*  
em\_esthum15:\*  
em\_esthum16:\*  
em\_esthum17:\*  
em\_esthum18:\*  
em\_esthum19:\*  
em\_esthum20:\*  
em\_esthum21:\*  
em\_esthum22:\*  
em\_esthum23:\*  
em\_esthum24:\*  
em\_esthum25:\*  
em\_esthum26:\*  
em\_esthum27:\*  
em\_esthum28:\*  
em\_estin1:\*  
em\_estin2:\*  
em\_estin3:\*  
em\_estin4:\*  
em\_estin5:\*  
em\_estom1:\*  
em\_estom2:\*  
em\_estov1:\*  
em\_estov2:\*  
em\_estpl1:\*  
em\_estpl2:\*  
em\_estpl3:\*  
em\_estpl4:\*  
em\_estpl5:\*  
em\_estpl6:\*  
em\_estpl7:\*  
em\_estpl8:\*  
em\_estpl9:\*  
em\_estpl10:\*  
em\_estrol1:\*  
em\_estrol2:\*  
em\_estrol3:\*  
em\_estrol4:\*  
em\_estrol5:\*  
em\_estrol6:\*  
em\_estrol7:\*  
em\_estrol8:\*  
em\_estrol9:\*  
em\_estrol10:\*  
em\_estrol11:\*  
em\_estrol12:\*  
em\_estrol13:\*  
em\_estrol14:\*  
em\_estrol15:\*  
em\_estrol16:\*  
em\_estrol17:\*  
em\_estrol18:\*  
em\_estrol19:\*  
em\_estrol20:\*  
em\_est25:\*  
em\_est26:\*  
em\_est27:\*  
em\_est28:\*  
em\_est29:\*  
em\_est30:\*  
em\_est31:\*  
em\_est32:\*  
em\_est33:\*  
em\_est34:\*  
em\_est35:\*  
em\_est36:\*  
em\_est37:\*  
em\_est38:\*  
em\_est39:\*  
em\_est40:\*  
em\_est41:\*  
em\_est42:\*  
em\_est43:\*  
em\_est44:\*  
em\_est45:\*  
em\_est46:\*  
em\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.







```

/db_xref="taxon:9606"
/clone="IMAGE:3603764"
/clone_lib="NIH_MGC_44"
/tissue_type="endometrium, adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      173 a      171 c      269 g      172 t
ORIGIN

      Query Match      4.1%; Score 36; DB 167; Length 785;
      Best Local Similarity 62.08; Pred. No. 6.2;
      Matches 57; Conservative 0; Mismatches 35; Indels 0; Gaps 0

Qy 604 gggcgctcgatgctctgtggggcgagcgctatccattgtccgcgtcgcggaaggt 663
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 214 GGGGCTTCCTGCTCCGTGGGGGCCCTGGCGGGGTATCCATGTTCGGTGGTGAGGT 273

Qy 664 ttbgtaggcaatcgcagggccttggttg 695
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 274 GCTGGAGAGGAGTTCGACAAAGCTTTGGTGG 305

RESULT 6
BF485823      650 bp      mRNA      EST      16-APR-2001
LOCUS      AT19875.5prime AT Drosophila melanogaster adult testes pOTB7
DEFINITION      Drosophila melanogaster cDNA clone AT19875 5 similar to CGI3768:
      FBAN0013768 last_updated:000321, mRNA sequence.
ACCESSION      BF485823
VERSION      BF485823.1 GI:11569124
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
      Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 650)
AUTHORS      Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman,
      B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan,
      D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,
      Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J.,
      Park, S., Paragas, V., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E.,
      Celniker, S. and Rubin, G.M.
      Berkeley Drosophila Gene Collection Project
      Unpublished (2000)
COMMENT      Contact: Stapleton, M.
      BDGP
      Lawrence Berkeley National Lab
      One Cyclotron Rd, Berkeley, CA 94720, USA
      Fax: 510 486 6798
      Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
      hit genomic sequence AE003614
      Plate: AT 198 row: G column: 3
      High quality sequence stop: 633.
      Location/Qualifiers
        1. .650
           /organism="Drosophila melanogaster"
           /db_xref="taxon:7227"
           /clone="AT19875"
           /clone_lib="AT Drosophila melanogaster adult testes pOTB7"
           /sex="male"
           /dev_stage="0-3 day old Ore-R males"
           /lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
           AT.121-AT.319: DH5-alpha Tona"
           /note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI;
           Site_2: XhoI; The mRNA for the testis library was made

```



cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center ([www.genome.clemson.edu](http://www.genome.clemson.edu)).

```

BASE COUNT      121 a   227 c   152 g   156 t      4 others
ORIGIN

Query Match      4.1%; Score 35.6; DB 238; Length 660;
Best Local Similarity 57.0%; Pred. No. 7.8;
Matches 65; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 529 gggaaactagattgggtgcgattcccgcttaccgttcggtcccaagatgtgtctcaa 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 549 GGGGAACAAGGTGTGTACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 490

Qy 589 gaccgtacactggacggcgctgatgtctctgtggggcgacggcggtatcc 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 489 GCGCGGACCAAGAAGCCGTCGATCGTCTCCGCGGCGGCGGCGGCGGCGGCGGCGG 436

RESULT 9
AZ127284/c      835 bp      DNA      GSS      02-JUN-2000
LOCUS
DEFINITION      OSJNB0084J06f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
                  clone OSJNB0084J06f, DNA sequence.
ACCESSION      AZ127284
VERSION      AZ127284.1 GI:8202809
KEYWORDS      GSS.
SOURCE      Oryza sativa.
ORGANISM      Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
AUTHORS      Wing, R.A. and Dean, R.A.
TITLE      A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL      Unpublished (1998)
COMMENT      Contact: Wing RA
                  Clemson University Genomics Institute
                  Clemson University
                  100 Jordan Hall, Clemson, SC 29634, USA
                  Tel: 864 656 7288
                  Fax: 864 656 4293
                  Email: rwing@clemson.edu
                  Seq primer: GTAAACGAGCGGCACGTG
                  Class: BAC ends
                  High quality sequence start: 34
                  High quality sequence stop: 745.
                  Location/Qualifiers
                    1..835
                      /organism="Oryza sativa"
                      /strain="Japanica"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:4530"
                      /clone="OSJNB0084J06f"
                      /clone_lib="CUGI Rice BAC Library (EcoRI)"
                      /tissue_type="Leaf"
                      /lab_host="E. coli DH10B"
                      /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrates.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and

```

Earle, 1991). The relatively small genome of rice, three times larger than that of *Arabidopsis*, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center ([www.genome.clemson.edu](http://www.genome.clemson.edu)).

```

BASE COUNT      133 a   287 c   202 g   210 t      3 others
ORIGIN

Query Match      4.1%; Score 35.6; DB 238; Length 835;
Best Local Similarity 57.0%; Pred. No. 8.3;
Matches 65; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 529 gggaaactagattgggtgcgattcccgcttaccgttcggtcccaagatgtgtctcaa 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 GGGGAACAAGGTGTGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 489

Qy 589 gaccgtacactggacggcgctgatgtctctgtggggcgacggcggtatcc 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 GCGCGGACCAAGAAGCCGTCGATCGTCTCCGCGGCGGCGGCGGCGGCGGCGGCGG 435

RESULT 10
BE605012/c      592 bp      mRNA      EST      21-AUG-2000
LOCUS
DEFINITION      WHE1713-1716_A23_A232S Wheat heat stressed spike cDNA library
                  Triticum aestivum CDNA clone WHE1713-1716_A23, mRNA sequence.
ACCESSION      BE605012
VERSION      BE605012.1 GI:9862282
KEYWORDS      EST.
SOURCE      bread wheat.
ORGANISM      Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Triticeae; Triticum.
REFERENCE      1 (bases 1 to 592)
AUTHORS      Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
                  , P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Malatrasi, M., Miller, R.,
                  Nguyen, H.T., Rausch, C.J., Seaton, C.L., Tong, J.C. and Zhang, D.
                  The structure and function of the expressed portion of the wheat
                  genomes - Heat stressed spike cDNA library
JOURNAL      Unpublished (2000)
COMMENT      Contact: Olin Anderson
                  US Department of Agriculture, Agriculture Research Service, Pacific
                  West Area, Western Regional Research Center
                  800 Buchanan Street, Albany, CA 94710, USA
                  Tel: 5105959773
                  Fax: 5105959818
                  Email: candernsnpw.usda.gov
                  Sequence have been trimmed to remove vector sequence and low
                  quality sequence with phred score less than 20
                  Seq primer: Stratagene SK primer.
                  Location/Qualifiers
                    1..592
                      /organism="Triticum aestivum"
                      /cultivar="Chinese Spring"
                      /db_xref="taxon:4565"
                      /clone="WHE1713-1716_A23_A23"
                      /clone_lib="Wheat heat stressed spike cDNA library"
                      /tissue_type="Whole spike"
                      /dev_stage="Spikes at 5, 10, 15 and 20 days after
anthesis"
                      /lab_host="E. coli SOLR"
                      /note="Vector: Lambda Uni-ZAP XR, excised phagemid;

```

using *E. coli* host strain XL1 Blue MRF'. Inserts were then subcloned by mass excision using ExAssist helper phage for conversion into phagemid vector pBK-CMV in *E. coli* host strain XL0LR. Sequenced using T3 primer: 5' ATT AAC CCT CAC TAA AGG GA 3'."

BASE COUNT	93 a	173 c	173 g	96 t	2 others
ORIGIN					
Query Match	4.0%; Score 35; DB 137; Length 537;				
Best Local Similarity	52.1%; Pred. No. 11;				
Matches 100; Conservative	0; Mismatches 91; Indels 1; Gaps 1;				

Db	434	CTTTGATGGCGCGCGGNCCTCCACGACGGGGATCTCGACACGAAAGATGCAGGCCGT	375
QY	595	gacctgacgggcgctcgatggtcctgtgtgggcgagcgcgctatccattgtcccgctcg	654
Db	374	GACCGAGCGCCCGCCT-GCTGCAGAGGCGGATGCCGGCTCGCGACGTGCCGCCCGCTGG	316
QY	655	gcggaaggttttggtaggcaattccgcgagccttggttggggactcttcccgaaacc	714
Db	315	CCAGTAGGTGCTCCACGAGCAGCAGCGCGTGCCTGTTGACGCGCTACGCCCTGGATGC	256
QY	715	caagctgctccc	726
Db	255	TCAGGTGGTCGC	244

LOCUS	CNS000E83	749 bp	DNA	GSS	04-JUN-1999							
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR28N12 of RPECI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.											
ACCESSION	AL069062											
VERSION	AL069062.1	GI:4949022										
KEYWORDS	GSS.											
SOURCE	fruit fly.											
ORGANISM	Drosophila melanogaster											
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyraoidea; Drosophilidae; Drosophila.											
REFERENCE	1 (bases 1 to 749)											
AUTHORS	Genoscope.											
TITLE	Direct Submission.											
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : RP 191 91006 EVRY cedex - FRANCE (E-mail : seoref@genoscope.cns.fr											

FEATURES	source	BASE COUNT
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial, PcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .	78 a	80 c
Location/Qualifiers	1..749	96 t
/organism="Drosophila melanogaster"		414 others
/db_xref="taxon:7227"		
/clone_lib="RPCI-98"		
/clone="BACR28N12"		
/note="end : T7"		



SOURCE Chlamydomonas reinhardtii.  
 ORGANISM Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 Chlamydomonadaceae; Chlamydomonas.  
 REFERENCE 1 (bases 1 to 529)  
 AUTHORS Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,  
 Nakamura,Y. and Tabata,S.  
 TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
 adapted cells of Chlamydomonas reinhardtii  
 JOURNAL DNA Res. 7 (5), 305-307 (2000)  
 MEDLINE 20539644  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.  
 FEATURES  
 source  
 1..529  
 /organism="Chlamydomonas reinhardtii"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone="ICL059f01\_r"  
 /clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; The cDNA library was constructed from cells cultured  
 in a carbon stress acclimatized condition in which carbon  
 dioxide concentration in the bubbling gas was changed from  
 5% to 0.04%".  
 BASE COUNT 103 a 169 c 163 g 93 t 1 others  
 ORIGIN  
 Query Match 4.0%; Score 34.8; DB 31; Length 529;  
 Best Local Similarity 50.0%; Pred. No. 13;  
 Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
 QY 415 ggagcggttaacccgtgaagctaacccgtggcggtggaagtagtagaacttggaacc 474  
 Db 405 GTACCAGATACCGCTCTGGCGGTTGCACTCGCGCGGTTGAAGCCGTAGTACTGTAGCC 346  
 QY 475 atgcgcacttggccttgaacccctcatctgcggatgcctacatggttgatggaaa 534  
 Db 345 CTCGAGACTTCTCGAAGCGCCCTCGTTCGTGTCGATGGCGGCGGAATGTTCTTGAA 286  
 QY 535 ctgattgggtcgatgcccgcttacgcttcgtcccaagaatgtgcttcaa 588  
 Db 285 CACGTTCCAGCGTACTTGAAGTGCTCGGACGCGACCAAGGGGTCTGTCCCA 232

Search completed: September 14, 2001, 06:59:27  
 Job time: 12074 sec







	5	RESULT	
		US-08-205697A-27/c	
		; Sequence 27, Application US/08205697A	
		; Patent No. 6218510	
		; GENERAL INFORMATION:	
		; APPLICANT: Sharpe, Arlene H.	
		; APPLICANT: Borriello, Francescopaulo	
		; APPLICANT: Freeman, Gordon J.	
		; APPLICANT: Nadler, Lee M.	
		; TITLE OF INVENTION: No. 6218510el Forms of T Cell Costimulatory Molecules	
		; AND USES THEREFOR	
		:	
Qy	13	caactggacatttgctcctaataatgatgaagcagcttcgaaggcgccctcttaagcc	72
Db	530	CAGCTTCGTAATTGGTTGCTCCCGGAGAGGCACGTCCGTGCCGCCGCCGCCGGC	471
Qy	73	ccttccaattccccctgcggtgagtgcagcgcttaaagctctcagagcatcacgtgggt	132
Db	470	GTCGATATTTCACGCCGCCGCCGCTCACGGCAGCATTCACGGCTCGAACAGCATCTGGC	411
Qy	133	cgagtgttgtgtatcgcgacc	153
Db	410	GTGTTTGTTCTTTCGAGCGCAGC	390

Query Match	4.0%	Score 35;	DB 3;	Length 38506;	Query Match	3.8%	Score 33;	DB 4;	Length 379;
Best Local Similarity	45.7%	Pred. No. 1.1;			Best Local Similarity	46.4%	Pred. No. 0.6;		
Matches 108;	Conservative	0;	Mismatches 125;	Indels 0;	Matches 108;	Conservative	0;	Mismatches 125;	Indels 0;
Gaps 0;					Gaps 0;				



```

RESULT          9
US-08-101-624-22/c
; Sequence 22, Application US/08101624
; Patent No. 5943607
; GENERAL INFORMATION:
; APPLICANT: Freeman, Gordon J.
; APPLICANT: Nadler, Lee M.
; APPLICANT: Gray, Gary S.
; TITLE OF INVENTION: No. 5942607el CTLA4/CD28 Ligands and
; TITLE OF INVENTION: Uses Therefor
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/101,624
; FILING DATE: 26-JUL-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```

```

RESULT 10
US-08-751-767A-5/c
. Sequence 5, Application US/08751767A
; Patent No. 5994104
; GENERAL INFORMATION:
; APPLICANT: ANDERSON, ROBERT J.
; APPLICANT: GRANT, HUGH
; APPLICANT: MACDONALD, IAN D.
; TITLE OF INVENTION: INTERLUKIN-12 FUSION PROTEIN
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,767A
; FILING DATE: 08-Nov-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 117-221
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164091
; TELEFAX: 7038164100
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 318..1181
US-08-751-767A-5

```

Query Match 3.8%; Score 33; DB 2; Length 1491;  
Best Local Similarity 46.4%; Pred. No. 1.1;  
Matches 108; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

	Query Match	3..8%	Score 33;	DB 2;	Length 1491;
	Best Local Similarity	46.4%;	Pred. No. 1.1;		
	Matches 108;	Conservative	0;	Mismatches 125;	Indels 0;
	Gaps				
QY	181	gaagtctctgtgcaagcagcgcggaatggtgtgtctgcaagcagcaaacctaagcgcaa	240		
Db	1442	GATGGCAGAAATGGAAACATGGCAAAAGAGAGGTTACATTAAAGCAANTGCATATGGGGA	1383		
QY	241	ctatctggagcctctgtgaaatcccggttaaccatcgccatcaacgcagattcgctatcc	300		
Db	1382	GAAAGGAGAGAGGATGCCAGCCATCTGTGAACGTGTTCAGCTTAATCTGAAAGAGGTTTTTA	1323		
QY	301	acatggttcttcccggtgttcaacgagtgactctctgggtggagcaacgcgtcacgctg	360		
Db	1322	AGNAGGTGAGGTAGTGGGTAAACACGCTTAATTTGTTAAGGGGAAGAAATGCCCATGATC	1263		
QY	361	cgcttggaagatgaagcgcaacacattactctgtcgggcggtggagatgtttt	413		
Db	1262	CCACGATCCATGTATCCCAAGAAGATGACGGAGGCTACCTTCAGATCTTTT	1210		

RESULT 11  
US-08-153-262-1/c  
; Sequence 1, Application US/08153262  
; Patent NO. 6071716  
; GENERAL INFORMATION:

```

; DATE: 15-OCT-1989
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1491
; US-08-153-262-1

Query Match 3.8%; Score 33; DB 3; Length 1491;
Best Local Similarity 46.4%; Pred. No. 1.1; Mismatches 125; Indels 0; Gaps 0;
Matches 108; Conservative 0;

QY 181 gaagtcttctgcaagcagcgcgagaaatggttctgcaagcagagaaactaaagcgcaa 240
DB 1442 GATGGCAGAGTGAACATCGGCAAAAGAGAGAGGTTACATTAAACCAATTCCATATGGGA 1383
QY 241 ctatctggacgcttctgtaaatcccgcttaaccatgccatcaacgcagattcgctatcc 300
DB 1382 GAAAGGAGAGGATGCGAGCCATCTTGTAACTGTTGAGCTTAATCTGAAAGAGGTTTTTA 1323
QY 301 acatggttctcccggttcaacgaggtagcttcttgggttgagcaaacgctcacgctg 360
DB 1322 AGAAGGTAGGTAGTGGGTAAACACGCTTAAATTTGTTAAGGGAAGAAATGCCCTCATGATC 1263
QY 361 cgttggagatgaagcgcacacattatctctgctgcgctggagatggtttt 413
DB 1262 CCCACGATCCATGATATCCGAGAAGAGATGACGGAGGCTACCTTCAGATCTTTT 1210

RESULT 12
US-08-479-744A-28/c
; Sequence 28, Application US/08479744A
; Patent No. 6084067
; GENERAL INFORMATION:
; APPLICANT: Freeman, Gordon J.
; APPLICANT: Nadler, Lee M.
; APPLICANT: Gray, Gary S.
; TITLE OF INVENTION: No. 6084067el CTLA4/CD28 Ligands and
; TITLE OF INVENTION: Uses Therefor
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSER: LAHIVE & COCKFIELD, LLP
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,744A
; FILING DATE: June 7, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/280,757
; FILING DATE: 26-JUL-1994
; APPLICATION NUMBER: 08/109,393
; FILING DATE: 28-AUG-1993
; APPLICATION NUMBER: 08/101,624
; FILING DATE: 26-JULY-1993
; APPLICATION NUMBER: 08/147,773
; FILING DATE: 3-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: RPI-004CP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1491 base pairs
; TYPE: nucleic acid

```



NOTES

STATE: Massachusetts  
COUNTRY: USA

—

```

; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/205,697A
; FILING DATE: 02-Mar-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: BWI-120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 318..1181
; US-08-205-697A-18

```

```

Query Match      3.8%; Score 33; DB 4; Length 1491;
Best Local Similarity 46.4%; Pred. No. 1.1;
Matches 108; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 181 gaagtcctgtgcaagcagcggaataatgggtgtgtgtgcaagcagcaaaactaaagcgcaa 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1442 GATGGCAGATGGAACATGGCCAAAGAGAGGTTACATTAAACCAATTGCCATATCGGGA 1383

QY 241 ctatctggagcgccttgcgtgaaatcccggttaaccatcgccatcaacgcagattcgctatcc 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1382 GAAAGGAGAGGGATGCCAGCATCTTTGTAACGTGTCAGCTTAATCTGAAAGAGGTTTTTA 1323

QY 301 acatgtttctcccgctgttcaacagaggtagcttcttggggggagcaacgcgtcacgctg 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1322 AGAAGGTGAGGTAGTGGGTAACACAGCTTAAATTTCTTAAGGGAAGAAATGCCCTCATGATC 1263

QY 361 cgcctggaagatgaagcgacacattatctctgctgcgcgctggagatgtttt 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1262 CCCACGATCCATGATATCCAGAGAGATGACGGAGGCTACCTTCAGATCTTTT 1210

```

Search completed: September 14, 2001, 07:51:47  
Job time: 12421 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2001, 03:41:38 ; Search time 2960.24 Seconds  
(without alignments)  
3715.094 Million cell updates/sec

Title: US-09-105-117i-1-copy\_1016\_1726  
Perfect score: 711  
Sequence: 1 ggtcgcgggcacggttgg.....cagattcgctatccacatgg 711

Scoring table: IDENTITY\_NUC  
Gapop 10\_0 , Gapext 1.0  
Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_vi:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_vil:\*  
59: gb\_vl2:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_rol:\*  
95: gb\_ro2:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	711	100.0	2374	2	CGLYSEG
C 2	711	100.0	2374	9	A93933
C 3	435	61.2	822	9	AX063771
C 4	406	57.1	993	9	AX063767
C 5	81.2	11.4	33285	3	SC5F8
C 6	71.8	10.1	12070	1	AB011413
C 7	63.2	8.9	14844	3	MBU34849
C 8	63.2	8.9	15239	10	I86263

```

9      62      8.7 10701 1 AE005522
10     62      8.7 26658 2 AP002563
11     61.6    8.7 38500 3 MTCY39
12     60.4    8.5 636   9 AX030085
13     60.4    8.5 8029 2 ECFDAPGK
14     60.4    8.5 10362 1 AE000375
15     60.4    8.5 14174 2 ECU28377
16     58.4    8.2 10542 1 AE004852
17     52.8    7.4 1804 2 ECRPIAA
18     52.8    7.4 1815 2 ECOICIA
19     52.8    7.4 6055 2 ECSERAICI
20     52.8    7.4 110000 66 AC020850_0
21     51.2    7.2 10367 1 AE005521
22     49.2    6.9 11168 1 AE004560
23     46.2    6.5 10572 1 AE004780
24     45.4    6.4 4513 10 AX067447
25     44      6.2 37218 3 MTCY2069
26     43.6    6.1 3632 1 AF037440
27     43      6.0 14789 1 AE004916
28     42.8    6.0 13193 1 AE005456
29     42.8    6.0 26857 2 AP002560
30     42      5.9 4532 3 SYOSRRA
31     41.4    5.8 4033 2 AS065741
32     40.6    5.7 10010 1 AE004639
33     40      5.6 10239 1 AE004457
34     40      5.6 183704 67 AC023024
35     39.8    5.6 2878 2 AHAYHYENE
36     39      5.5 160938 77 AC090056
37     38.8    5.5 1033 1 AB053349
38     38.8    5.5 4217 2 AY017305
39     38.6    5.4 1070 3 PSECATRA
40     37.8    5.3 11015 1 AE004736
41     37.6    5.3 5644 1 AF061070
42     37.2    5.2 10614 1 AE004351
43     37.2    5.2 187111 88 AC087244
44     37      5.2 1592 3 S80112
45     37      5.2 1790 3 PJPTEFDRII

```

## ALIGNMENTS

```

RESULT 1
CGLYSEG 2374 bp DNA
LOCUS C.glutamyl lysE and lysG genes.
DEFINITION X96471
ACCESSION X96471
VERSION X96471.1 GI:1729753
KEYWORDS lysE gene; lysG gene; Lysine export regulator protein; Lysine
exporter protein; Lysine governor.
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 2374)
AUTHORS Vrljic.M., Sahn.H. and Eggeling,L.
TITLE A new type of transporter with a new type of cellular function:
L-lysine export from Corynebacterium glutamicum
JOURNAL Mol. Microbiol. 22 (5), 815-826 (1996)
MEDLINE 97126810
REFERENCE 2 (bases 1 to 2374)
AUTHORS Vrljic.M.M.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-1996) M.M. Vrljic, Institut fuer Biotechnologie
1, Forschungszentrum Juelich, Postfach 1913, D-52425 Juelich, FRG
FEATURES
source
location/Qualifiers
1..2374
/organism="Corynebacterium glutamicum"
/strain="R127"
/db_xref="taxon:1718"
complement(82..954)
gene
/gene="lysG"

```

```

CDS
complement(82..954)
/gene="lysG"
/note="lysine governor"
/codon_start=1
/transl_table=11
/product="lysine export regulator protein"
/protein_id="CAA65323.1"
/db_xref="GI:1729754"
/db_xref="SWISS-PROT:P94632"
/translation="MNPQLDTLLSIIDEGSGFEGASLALSTSPSAVSORVRALEHHVG
RVLSRTQPAKATEAGVILVQARKMVLQAETKAQLSRLAEIPLTIAINADLSUTW
FPFVNEVASMGATLTLEDEAHTLSLLRRGDVLGAVTREAANVACVEVGLTMR
HLAIATPSLDAYMVDGLDWAAMPVLRFGKDVLDQDLQDGRVDPGVRGRRVSTPS
AEGFGEAIRRGGLGWLLETPQAPMLKAGEVILLDEIPDTPMYQWRLESRLARL
TDAVYDAIEGLRP"
1025..1726
/gene="lysE"
1025..1726
/gene="lysE"
/codon_start=1
/transl_table=11
/product="Lysine exporter protein"
/protein_id="CAA65324.2"
/db_xref="GI:13397387"
/db_xref="SWISS-PROT:P94633"
/translation="MEIFITGLLGASLLSTGPONVILVIRKQIREGLIALLVCLLI
SDVFLFIATGLVDLLSNAAPITVILDMRGGTAYLLWFAVMAAKDAMTKVAPQIIE
ETEPVDPDTPLGGSATVDPRNVRVSVSDKORVVKPMLMALVLTWLPNAYLDA
FVFVGGVAGVGDYGRWTFAPAGAFASLIWFLPLVGFGAALSRPLSSPKVRWVNVV
AVMTALAIKLMVG"
complement(1723..2352)
/gene="orf3"
complement(1723..2352)
/gene="orf3"
/codon_start=1
/transl_table=11
/protein_id="CAA65325.1"
/db_xref="GI:1729756"
/db_xref="SPTREMBL:P94634"
/translation="MYALRDIVASGKALYVGISSYSGPELTAAEAFAEAEPCPLLIHQ
PSYLIINRVEEPDGDENLLQSAANGLGVIAFSPLAGLLTDKYLDTIGESRSASQ
GKLSSEGMNVNNDIMVRKLDIAQERQSIAQMALAVLRQEGYIGADTVTSALIGA
SSVQDLSLDLSLNNLEFSDAELEAIDISHDGINIWAATDSKTREN"
BASE COUNT 526 a 640 c 648 g 560 t
ORIGIN
Query Match 100.0%; Score 711; DB 2; Length 2374;
Best Local Similarity 100.0%; Pred. No. 2.7e-207;
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtgtatcgggcacggttgcttcttcaatgatctgtggcgcttccacctgttt 60
|||||
Db 1359 GTGTATCGGGCACGGTGTGTTCTTCTTCAATGATCTGTGGCGCTTCCACCTGTTT 1300
QY 61 gtcatggcgctttctcgctccatgacggcaaacaccataacaggtgaagcgatgcacccag 120
|||||
Db 1299 GTCATGGCGCTCTTCGCTCCATGACGGCAACACCAACAGGTAAGCGATGCCACCCAG 1240
QY 121 cgcataataatcagacgatcgcgcgcatgggacaaagataacacgcaaggtgcg 180
|||||
Db 1239 CGCATAAATATCAGCACGATCGCGCGCATTTGGACAAAAGATCAACGCCCAAGGTGCGG 1180
QY 181 gcgatgaacaaagacgctcagaataataacacacgagagacacgcaatgagtccttcg 240
|||||
Db 1179 GCGATGAACAAAAGACGTCAGAAATTAAACACACGAGAACCCCATGAGTCTCTCG 1120
QY 241 cgtctaattccttgtttaataaccagtagtacattctgcggtccgtaggacagtaaaagactg 300
|||||
Db 1119 CGCTTAATTCCTTGTGTTAATCACCAGTACATTCTCGGTCCGATGACAGTAAAGACTG 1060
QY 301 gcccccaagcagacctgtaataagaatttcattccattgaccattgacattggaagt 360
|||||

```

```
Db 1059 GCCCCAAGACAGACCTGTAATGAGATTTCATGATCACCATCGTGACCTATGGAAGT 1000
Qy 361 acttaagtaaaatgattggttcttaacatggtttaaataatagcttcaatgaacccattcaa 420
Db 999 ACTTAAGTAAATGATTGTTCTTAACATGTTTAAATATAGCTTCATGAACCCCAATTCAA 940
Qy 421 ctgacacattgctctcaatcatgatgaagcagcttcgaagcgccctcttagccctt 480
Db 939 CTGGACATTTGCTCTCAATCATGATGAAGCAGCTTCGAAGGCGCTCCTTAGCCCTT 880
Qy 481 tccatttccccctcggggtgagtcagcgctttaaagctctcgcagcatcacagtggtcgca 540
Db 879 TCCATTTCCCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGACATCAGTGGTGA 820
Qy 541 gtgttgatcgccacccacccgcccgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 600
Db 819 GTGTTGGTATCGCGCACCAACCGCCAAAGCAACCGAAGCGGTGAAGTCTTGTGCAA 760
Qy 601 gcagcgagaaatggttctgctgaagcagaactaaagcgaactatctggagccctt 660
Db 759 GCAGCGGGAATGTTGCTGCAAGCAGAACTAAAGCGCAACTATCTGCGACGCTT 700
Qy 661 gctgaatcccgcttaaccatcgccatcaacgagattcgctatccacatgg 711
Db 699 CCGAATCCGTTAACCATCGCCATCAACGACATTCGCTATCCACATGG 649

RESULT 2
A93933/c
LOCUS A93933 2374 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 2 from Patent WO9723597.
ACCESSION A93933
VERSION A93933.1 GI:6742037
KEYWORDS
SOURCE
ORGANISM Corynebacterium glutamicum.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE
1 (bases 1 to 2374)
AUTHORS Vrljic, M. and Eggeling, L.
TITLE PROCESS FOR THE MICROBIAL PRODUCTION OF AMINO ACIDS BY BOOSTED
ACTIVITY OF EXPORT CARRIERS
JOURNAL Patent: WO 9723597-A 2 03-JUL-1997;
KERNFORSCHUNGSANLAGE JUELICH (DE); VRLJIC MARINA (DE)
FEATURES
Location/Qualifiers
source 1..2374
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
BASE COUNT 526 a 640 c 648 g 560 t
ORIGIN

Query Match 100.0%; Score 711; DB 9; Length 2374;
Best Local Similarity 100.0%; Pred. No. 2.7e-207;
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgtcatcgagcaggttggttcttctcaatgatctgtggttcacattgttt 60
Db 1359 GTGTCATCGGCGGTTGTTCTGTTCTTCATGATCTGTGCGCTTCCACCTTGT 1300
Qy 61 gtcattgctcttctgctgcacatgacggcgaacaccataacagggttaagcgatgccacccag 120
Db 1299 GTCATGCGCTTTCGCTGTCATGACGCGCAACCAACAGGTAAAGCGATGCCACCCAG 1240
Qy 121 cgcataatctcgacagcatcgccggttcattgacaaaagatacaagcccaaggtaccg 180
Db 1239 CGCATATATCGACCATCGCGCGCGCATTTGACAAAAGATCAACGCCAAGGTGCG 1180
Qy 181 gcgatgaacaaaagcgtcagaaattaaacacacagagaagaccgcaatgagtcctcg 240
Db 1179 GCGATGAACAAAAGCGTCAGAAATTAACACACGAGAGAACCGCAATGAGTCTTCG 1120
```

```
Qy 241 ogcttaattcttbttaataccacagtacattctcggtccgatggacagtaaaagactg 300
Db 1119 CGCTTAATTCCTTTTAAATCACCAGTACATTCGCGTCCGATGGAGTAAAGAACTG 1060
Qy 301 gccccaaaagcagacacgttaatagaatttccatgatcacccatcgtagcctatggaagt 360
Db 1059 GCCCCAAGACAGACCTGTAATGAGATTTCATGATCACCATCGTGACCTATGGAAGT 1000
Qy 361 acttaagtaaaatgattggttcttaacatggtttaaataatagcttcaatgaacccattcaa 420
Db 999 ACTTAAGTAAATGATTGTTCTTAACATGTTTAAATATAGCTTCATGAACCCCAATTCAA 940
Qy 421 ctgacacattgctctcaatcatgatgaagcagcttcgaagcgccctcttagccctt 480
Db 939 CTGGACATTTGCTCTCAATCATGATGAAGCAGCTTCGAAGGCGCTCCTTAGCCCTT 880
Qy 481 tccatttccccctcggggtgagtcagcgctttaaagctctcgcagcatcacagtggtcgca 540
Db 879 TCCATTTCCCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGACATCAGTGGTGA 820
Qy 541 gtgttgatcgccacccacccgcccgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 600
Db 819 GTGTTGGTATCGCGCACCAACCGCCAAAGCAACCGAAGCGGTGAAGTCTTGTGCAA 760
Qy 601 gcagcgagaaatggttctgctgaagcagaactaaagcgaactatctggagccctt 660
Db 759 GCAGCGGGAATGTTGCTGCAAGCAGAACTAAAGCGCAACTATCTGCGACGCTT 700
Qy 661 gctgaatcccgcttcaaccatcgccatcaacgagattcgctatccacatgg 711
Db 699 CCGAATCCGTTAACCATCGCCATCAACGACATTCGCTATCCACATGG 649

RESULT 3
AX063771/c
LOCUS AX063771 822 bp DNA PAT 24-JAN-2001
DEFINITION Sequence 53 from Patent WO0100843.
ACCESSION AX063771
VERSION AX063771.1 GI:12541483
KEYWORDS
SOURCE
ORGANISM Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE
1 (bases 1 to 822)
AUTHORS Pompeius, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
TITLE corynebacterium glutamicum genes encoding metabolic pathway
JOURNAL Patent: WO 0100843-A 53 04-JAN-2001;
BASF AKTIENGESSELLSCHAFT (DE)
FEATURES
Location/Qualifiers
source 1..822
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
CDS
101..802
/notes="RXA01394"
/codon_start=1
/transl_table=11
/protein_id="CAC25127.1"
/db_xref="GI:12541484"
BASE COUNT 167 a 192 c 246 g 217 t
ORIGIN

Query Match 61.2%; Score 435; DB 9; Length 822;
Best Local Similarity 100.0%; Pred. No. 1.6e-122;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 ggtgcatcgccgacggtgtgttcttcaatgatctgtggccttccacctgttt 60  
 DB 435 GTGTCTATCGGGACAGGTTGGTCTGTCTTCAATGATCTGTGGCGCTTCCACCTTGT 376  
 QY 61 gtcattggtcttcttctgctgacgagcggcaaacataaacaggttaagcgtatgcaccccccag 120  
 DB 375 GTCATGGCGTCTTCGCTGCCATGACGGCAAAACCATACAGGTAAAGCGATGCCACCCAG 316  
 QY 121 cgcataatcagacgacgtatcgccggtgattggaacaaagatcaacgcccaggtgcg 180  
 DB 315 CGCATTAATATCAGACAGCATCGCGCGGATTTGGACAAAAGATCAACGCCCAAGGTGCGG 256  
 QY 181 gcgatgaacaaagacgctcagaataataacacacagagaaacgcgaatgagtcctcg 240  
 DB 255 GCGATGAACAAAAGACGTCAGAAATTAACACACAGAGAAACCCCAATGAGTCTTCG 196  
 QY 241 cgttaattccttcttgaatcaccagatcattctgcgttcggtatgacagtgaaagactg 300  
 DB 195 CGCTTAATTCCTTGTAAATCACCACTACATCTTCCGCTCCGATGACAGTAAAGACTG 136  
 QY 301 gccccaaagacagactgaatgaagatttccatgatcaccatcgacccatggaagt 360  
 DB 135 GCCCCAAAGACAGCTGAATGAAGATTTCCATGATCACCATCGTACCTATGGAAGT 76  
 QY 361 acttaagtaaatgattgttcttcaatcaggttttaatatagcttccatgaacccattcaa 420  
 DB 75 ACTTAAGTAAATGATTGTTCTTAACATGTTTATATATAGTTCATGAACCCCATCAA 16  
 QY 421 ctggacactttgtc 435  
 DB 15 CTGGACACTTTGCTC 1

## RESULT 4

LOCUS AX063767 993 bp DNA PAT 24-JAN-2001  
 DEFINITION Sequence 49 from Patent WO0100843.  
 ACCESSION AX063767  
 VERSION AX063767.1 GI:12541479  
 KEYWORDS Corynebacterium glutamicum.  
 SOURCE Corynebacterium glutamicum.  
 ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 REFERENCE 1 (bases 1 to 993)  
 AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberbauer, G.  
 TITLE Corynebacterium glutamicum genes encoding metabolic pathway proteins  
 JOURNAL Patent: WO 0100843-A 49 04-JAN-2001;  
 BASF AKTIENGESSELLSCHAFT (DE)  
 FEATURES  
 Source 1..993  
 /organism="Corynebacterium glutamicum"  
 /db\_xref="taxon:1718"  
 CDS 101..973  
 /note="RXA01393"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAC25125.1"  
 /db\_xref="GI:12541480"  
 /translation="MNPQLDTLLSIIDGSEFAGSLAISPSAVSRVRALEHHVG  
 RLVSRTPAKATAGEVLVQAAKRWLLQAEKALSGRLAEIPLTAINADSLSTW  
 RYVNEVASGGATLRLLEDAHTLSLRGDLVLTAVREANVAGCEVVELTMR  
 HLAITPSLRDAYVDGKLDWAMPVLRFGPKDVLQDRDLGRVDGPGVRRRVSTPS  
 AEGFGEARLRGGLGELLPEPTQAPMLKAGEVILLDEIPDTPMYQWRLESLRL  
 TDVVDAAEIGLRP"  
 BASE COUNT 233 a 256 c 268 g 236 t  
 ORIGIN

Query Match 57.1%; Score 406; DB 9; Length 993;

Best Local Similarity 100.0%; Pred. No. 1.4e-113;  
 Matches 406; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 caaagcagacacgttaataagaatttccatgatcaccatcgtgacatcgtgaagtactta 365  
 DB 1 CAAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTGACATATGGAAGTACTTA -60  
 QY 366 agtaaatatgattggttcttaacatggtttaaataatagcttcaaccccatcaactga 425  
 DB 61 AGTAAATGATTGGTTCCTTAACATGTTTAAATATAGCTTCATGAACCCCATCAACTGGA 120  
 QY 426 cactttgctctcaatcattgatgaagcgagcttcgaagcgctcctccttagcctttccat 485  
 DB 121 CACTTTGCTCTCAATCATTTGATGAAGGCAGCTTCGAAGGCGCTCTTAGCCCTTTCCAT 180  
 QY 486 ttccctcctcgccggtgagtcagcggttaaaagctctcgaagatcaactggtggtcagtt 545  
 DB 181 TTCCCTCCTCGGGGTGAGTCAGCGGCTTAAAGCTCTCGAGCATCATCGTGGGTGAGTGT 240  
 QY 546 ggtatcgccaccccaaccgcccgaagcaaccgaagcggtgaagtctcttgcaagcagc 605  
 DB 241 GGTATCGCGCACCCCAACCGGCCCAAGCAACGAGGGGTGAAGTCTTGTGCAACGACG 300  
 QY 606 gcggaataatggtgtgtctgcaagcgaataaaagcgaactatctggaagcgttctga 665  
 DB 301 GCGGAATAATGTTGTTGCTGCAAGCAGAAACTAAAGCGCAACTATCTGGACGGCTTCTGA 360  
 QY 666 aatcccggttaacatcgccatcaacgagatctcgtatccacatgg 711  
 DB 361 AATCCGGTTAAACATCGCCATCAACGACAGATTGCGTATCCACATGG 406

## RESULT 5

LOCUS SC5F8 33285 bp DNA BCT 09-JUN-2000  
 DEFINITION Streptomyces coelicolor cosmid 5F8.  
 ACCESSION AL357613  
 VERSION AL357613.1 GI:8347023  
 KEYWORDS amino acid decarboxylase; anti-sigma factor; integral membrane protein; LuxR-family regulator; LysR-family transcriptional regulator; membrane transport protein; oxidoreductase; regulatory protein; RNA polymerase sigma factor; serine/threonine protein kinase; stress-inducible protein; TetR-family transcriptional regulator; thioredoxin reductase; TetR-family transcriptional two-component sensor histidine kinase.  
 SOURCE Streptomyces coelicolor A3(2).  
 ORGANISM Streptomyces coelicolor A3(2).  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 REFERENCE 1 (bases 1 to 33285)  
 AUTHORS Redenbach, M., Kieser, H.M., Denapaite, D., Eichner, A., Cullum, J., Kinashi, H. and Hopwood, D.A.  
 TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
 JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)  
 MEDLINE 97000351  
 REFERENCE 2 (bases 1 to 33285)  
 AUTHORS Seeger, K.J. and Harris, D.  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 33285)  
 AUTHORS Thomson, N.R., Parkhill, J., Barrell, B.G. and Rajandream, M.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-JUN-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: wellcome@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK  
 COMMENT Notes:  
 Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))

CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1994) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>.  
CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
Cosmid 5F8 lies on genomic restriction fragment AseI-A bordered by cosmids 5H1 and 4G10.

#### FEATURES

source  
1. .33285  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid 5F8"  
complement(1. .344)  
/gene="SC5F8.01c"  
complement(1. .108)  
/gene="SC5F8.01c"  
/note="nominal overlap with cosmid SC5H1 between bases 1. .108."  
CDS  
complement(1. .344)  
/gene="SC5F8.01c"  
/note="SC5F8.01c, possible serine/threonine protein kinase, partials CDS, len: 114 aa. Similar to many proposed Streptomyces coelicolor serine/threonine protein kinases e.g. TR:CAB82014(EMBL:AL161755) SC63.07 (717 aa), fasta scores opt: 172 z-score: 222.0 E(): 6.8e-05 33.0% identity in 103 aa overlap. Overlaps and extends into CDS SC5H1.01 on the adjoining cosmid."  
/codon\_start=1  
/transl\_table=11  
/product="putative serine/threonine protein kinase (fragment)."  
/protein\_id="CAB93729.1"  
/db\_xref="GI:8347024"  
/translation="MGRHVSHELVAGRYLFEVQVRETNRVCHSGEDATGRCPLV TRIEPEGRAGGAARAPGRVITGETMASLCPGRIAPVLDVADGMLWTVEWVAG VPLGDLDDRGA"  
complement(492. .1451)  
/gene="SC5F8.02c"  
complement(492. .1451)  
/gene="SC5F8.02c"  
/note="SC5F8.02c, possible threonine dehydratase, len: 319 aa. Similar to many including: Escherichia coli SW:THD2.FCOLI(EMBL:M21312) threonine dehydratase catabolic (EC 4.2.1.16) TdcB (329 aa), fasta scores opt: 609 z-score: 676.6 E(): 3.2e-30 39.4% identity in 312 aa overlap. Contains a Prosite hit to Serine/threonine dehydratases pyridoxal-phosphate attachment site and a pram match to entry PF00291 S\_T\_dehydratase, Pyridoxal-phosphate dependent enzyme."  
/codon\_start=1  
/transl\_table=11  
/product="putative threonine dehydratase."

/protein\_id="CAB93730.1"  
/db\_xref="GI:8347025"  
/translation="MIAITEIAAAERIAGHVVRTPTVPSPGLSALLGVPVTKLELL QRTSGFKARGATAKLISLTERAGACVAVSGNGHGIJAVVMAADLVKATVVMPTA PARSVIAEAGALVRLTDMDSAPALVRLREEGILTVPDFDPVAVAGQCTVGLGF AEDASLTDVLSIGGGGLIAGVAALRALRPGVRVWGVEYETGAEMSAALAGSPLT CLPSSVVTLSAFPSVSRILYDHVAELVTELVVPDREAVRGSLALAEHAKVWTEPAAG CILPAARRVVERGVGRIGLVVCGGNATVGDMAVWADRFLGR"  
complement(546. .1412)  
/gene="SC5F8.02c"  
/note="Pfam match to entry PF00291 S\_T\_dehydratase, pyridoxal-phosphate dependent enzyme, score 210.10, E-value 3.3e-59"  
E-value 3.3e-59  
complement(1287. .1328)  
/gene="SC5F8.02c"  
/note="PS00165 Serine/threonine dehydratases pyridoxal-phosphate attachment site"  
complement(1477. .2154)  
/gene="SC5F8.03c"  
complement(1477. .2154)  
/gene="SC5F8.03c"  
/note="SC5F8.03c, unknown, len: 225 aa. Similar to several proteins of undefined function including: Deinococcus radiodurans TR:Q9RWP1(EMBL:A5001920) conserved hypothetical protein (254 aa), fasta scores opt: 467 z-score: 555.6 E(): 1.8e-23 36.8% identity in 205 aa overlap and Streptomyces coelicolor TR:CAB70638(EMBL:AL137242) SC8F4.12C (239 aa), fasta scores opt: 397 z-score: 474.2 E(): 6.1e-19 36.1% identity in 183 aa overlap."  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC5F8.03c."  
/protein\_id="CAB93731.1"  
/db\_xref="GI:8347026"  
/translation="MKGDLFSSEHMVQFATPGMTVENSKCIRYAVNGEMLARQCAMI ATRGNLQFERKQGVGMUKRAVTEGLEPLMAVRQGGENAFANEAQCFVVEFGDE FTVNGRNVLDCASISYRIATVKGSIAGGLFNSTVFTGGLVLCENPLVIPVSQ QYPIHVDTDVAVVGVSAGATSLHRSQISGMSLGGSGEAVQLVLQGGFVVVVRPEAT POKPQOH"  
complement(2166. .3470)  
/gene="SC5F8.04c"  
complement(2166. .3470)  
/gene="SC5F8.04c"  
/note="SC5F8.04c, unknown, len: 434 aa. Similar to Streptomyces coelicolor TR:Q9XW9(EMBL:AL096743) putative transport associated protein, SC17.21 (399 aa), fasta scores opt: 428 z-score: 449.2 E(): 1.5e-17 38.3% identity in 209 aa overlap."  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC5F8.04c."  
/protein\_id="CAB93732.1"  
/db\_xref="GI:8347027"  
/translation="MSDLVPGGNVPLPGGVSVRVPGGFDVSLVTDGGKVGVDADF VTYNPEAPGARLDRDITLVDPARLRGAARVTVAVGSDPGTPIGALPSPVLVTD A RGRTVARTPARPREGTLLVLFYFRRGGKWRALGOGYADGLAGLDFGVEYTD D APPEPATPELTPGADSDGFLNVNSRAAAGSPAVRDLRAAHAAHMAAAG TLSIETRDGVSQHVQVSGAFYLVVGEHLVSGPRTPAEFVAYCLRAERTRLHDTA FTHAGWACVTGPGSDTYWTALWAVPLTPDGLARTTAEVVGLTNRERAGLUPALVD ARLTAAQAHSADMYTRDFYSHTDPDGCKPWRRAAGADRESVGENIACGQSPAEV VEGMNSPGHRAINLEAGFTHIGVLAGGGRAGTWTQLLG"  
complement(3562. .6528)  
/gene="SC5F8.05c"  
complement(3562. .6528)  
/gene="SC5F8.05c"  
/note="SC5F8.05c, possible LuxR-family regulator, len: 988 aa. Similar to several other putative regulators from Streptomyces coelicolor including: TR:Q9XAN4(EMBL:AL079355) SC4C6.06 (893 aa), fasta scores opt: 305 z-score: 328.7 E(): 7.7e-11 26.4% identity in 910 aa overlap and TR:Q92573(EMBL:AL035569) SC8D9.18 (1091 aa), fasta scores opt: 857 z-score: 926.1 E(): 0.36.5%





```

TITLE      Virulence-attenuating genetic deletions deleted from mycobacterium
JOURNAL    BCG
PATENT     Patent: US 5700683-A 17 23-DEC-1997;
FEATURES   Location/Qualifiers
Source     1..15239
           /organism="unknown"
BASE COUNT 2878 a 4791 c 4729 g 2841 t
ORIGIN

Query Match      8.9%; Score 63.2; DB 10; Length 15239;
Best Local Similarity 55.3%; Pred. No. 6.le-08;
Matches 166; Conservative 0; Mismatches 128; Indels 6; Gaps 2;

QY 418 caactggacacttgctctcaatcattgatgaaggcagcttcgaagcgccctccttagcc 477
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5253 CAGCTGGCGCATGGCTGCGTGGTCGAACATGGGAGCGCTTCATGCGCGCGGAGCGC 5312

QY 478 ctttcatttccccctcgcggtgagtcagcggttaaaagctctcagacatcacgtgggt 537
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5313 CTACATGTCACCCCTCGCTGTCAGTCAGCGCATCAAGTCGTTGGAGCAGCGGTCGCG 5372

QY 538 cgaagtgttgatcgccaccacacgcgccaagaacgaacgagcggtggaactccttg 597
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5373 CAGGTCTGCTGGTCTAGGGAAAAGCATCTCGGGCGCAGCAGCGCATCCGCTGTTG 5432

QY 598 caagcagcgcgaaatggtgtctgcaagcagcaaaactaaagcgcaactctctg--ac 655
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5433 CGGTTGCGCGCGCAACACAGCGTGTCTGAGTCGAGGCGCGCTGCTGAATGGTGGCAAC 5492

QY 656 gcccttgctgaa----atcccggttaaccatcgccatcaacgacgactcgtctccacatgg 711
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5493 GGCTCGCTGAACGCGACCGGGATCACCATTGCGGTAAACCGCGCATCCATCGCGACATGG 5552

RESULT      9
AE005522    AE005522 10701 bp DNA BCT 21-MAR-2001
LOCUS       Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 141
DEFINITION of 290.
ACCESSION   AE005522 AE005174
VERSION     AE005522.1 GI:12517448
KEYWORDS    Escherichia coli O157:H7 EDL933.
SOURCE      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 10701)
AUTHORS    Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Postai,G., Hackett,J., Link,S., Boutin,A., Shao,Y., Miller,L.,
            Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
            Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
            Nature 409 (6819), 529-533 (2001)
            21074935
            11206551
REFERENCE   2 (bases 1 to 10701)
AUTHORS    Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Postai,G., Hackett,J., Link,S., Boutin,A., Shao,Y., Miller,L.,
            Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
            Direct Submission
            Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
            Location/Qualifiers
            1..10701
            /organism="Escherichia coli O157:H7 EDL933"
            /strain="EDL933"
            /serotype="O157:H7"

```

```

/db_xref="taxon:155864"
/note="enterohaemorrhagic"
133..2277
/gene="sbm"
/note="24254"
133..2277
/gene="sbm"
/function="enzyme; Degradation of small molecules: Fatty
acids"
/note="Residues 1 to 714 of 714 are 99.15 pct identical to
residues 1 to 714 of 714 from Escherichia coli K-12 Strain
MG1655; B2917"
/codon_start=1
/transl_table=11
/product="methylmalonyl-CoA mutase (MCM)"
/protein_id="AAG58043.1"
/db_xref="GI:12517449"
/translation="MSNEQEWQOLANKELSRREKTVDSLVQQTAEGLIAIKPLYTEADL
DNLEVTGLPGLPYPRVAGTMTAQPTIRQVAGFTAKESNAFYRNLAAGOKGL
SVAPDLATHRGVSDNPRVAGDVGKTVAIDTDMKVLFDQIPDKMSWMTMNGAV
LPVLAFYIAAEQGVGPKLTGTIONDILKEYLCRNTYIYPPKSMRIADIIAMCS
GNMPFWTISISGYHMGAGANCVOAFTIADGIEYIKAAISAGLKIDDPAPLSFF
FGIGMDLFMNVAMLRARYLWSEAVSGGGAQPKSLADKTHCQTSGWSLTCQDPNNV
IRTTTEALATLGGTQSUHTNAFDEALGLTDFDSARIARTQIILQIESELCRTVDPL
AGSYIESITDQIQVQARAIQQIDEAGGMAKIEAGLPKRMIEESAREQSLIDQK
RVYGVNKKYKLDHEDETDVLIEDNVNVEQIASLERIRATRDAAVTAALNATHAA
QHNNLLAAVNAARVRATLGEISDALEAFDRYLVPSOCVTGIVAAQSYHOSEKSAE
FDAVAOPEFLADNGRPRILIAKMGLDGHDGAKVTASAVSDLGFDVDSLPMFSTP
BEIARLAVENDYHVVGASSLAAGHKTLIPELVEALKKWGREDCICWAGGVIPPDYAF
LOERGVAAIYGGTGMPLDSVRDVLNLISQHH"
2270..3265
/gene="ygfD"
/note="24255"
2270..3265
/gene="ygfD"
/function="phenotype; Not classified"
/note="Residues 1 to 331 of 331 are 99.69 pct identical to
residues 1 to 331 of 331 from Escherichia coli K-12 Strain
MG1655; B2918"
/codon_start=1
/transl_table=11
/product="putative nucleotide-binding protein"
/protein_id="AAG58044.1"
/db_xref="GI:12517450"
/translation="MINEATLAESIRRLRQGEHATLAQAMTLVSRHPRHQAALSTQLL
DAIMPYCGNTLRIGVTGTPGAGKSTFLFAFGMLLIREGLKVAIVADPSPVTVGSIL
GDKTRMDLARAFAEAFIRPVSSGHILGCASORARELMILCEAAGVDVIVETVGVGYS
ETEVANWVDCFISIQIAGGDDDLQIGIKKGLMEVADLVINKDDGDNHNTVAIRHMYE
SALHILRRKYDEWPVRVLTCSALEKRGIDEIWHALIDFKALTASGRQQVQQQSVSE
WLRKQTEEEVLNHLFANEDFDYRYQTLLAVKNNTLSPTGLQLSEFIQOYFD"
3234..4061
/gene="ygfG"
/note="24256"
3234..4061
/gene="ygfG"
/function="putative enzyme; Not classified"
/note="Residues 1 to 275 of 275 are 100.00 pct identical
to residues 1 to 275 of 275 from Escherichia coli K-12
Strain MG1655; B2919"
/codon_start=1
/transl_table=11
/product="putative enzyme"
/protein_id="AAG58045.1"
/db_xref="GI:12517451"
/translation="MNLSSRRNLIKIFMSYQYVNVTVINKVAVFTEFNYGRKLNALS
VFIDDLQALSDLNRPETRCILRAPSGSKVFSAGHDHHELPGSGRDPVSLVDDPLROI
TRLOKQPKPIISWEGSVGCAFEIMSSDLIIIAASTSTFSTFMTPNVNLGVPNLVGHI
NLTRDAGFHVIKELIFTASPIAQRALAVGLNHVVEVEEDETTLQMAHHISEKAPL
AIAVTKELRVLGEAHTMNSDEFRIQGRRAVYDSEYQEGMNAFLFKRKPNEVGH"
4085..5563
/gene="ygfH"
/note="24257"
4085..5563

```



```

/gene="ygfH"
/function="putative enzyme; Not classified"
/notes="Residues 1 to 492 of 492 are 99.39 pct identical to
residues 1 to 492 of 492 from Escherichia coli K-12 Strain
MG1655: B2920"
/codon_start=1
/transl_table=11
/product="putative coenzyme A transferase"
/protein_id="AAG58046.1"
/db_xref="GI:12517452"
/translation="METQWTRMTADAEAEIIOHNDVAFSGFTFPGSPKALPTAIARR
ANQEHAKPKYQIURLITGASISAAADDDVLSADAVSWRAPYOTSSGLKKINGAVSF
VDLHSEVQAVNVPFGDIDVAVIASALAPGRVLTSGIGNAPTLLRAKKVILE
LNHIDPRVLAEDIVIPGAPRRNSVIFHANDRVGTRIVQIDPKIVAVVETNLPD
AGNMLDKQPMCOQIADNVVTFLLQEMAGHRIPPEFLQSGVGVNINNAVMARLGNP
ETPPFMYSEVLOQSVHLETKIGSASASLTIISADSLRKIYDNMDYFASRIYLRP
QETSNNPEIIRLGLVLANVLEFDIYGHANSHVAGVLMNGIGSGDFPNAVYLSI
FMAPSATKEGISTVVPKMSHVNDHNSVKVITTEOGIADLRGLSLPLQARTIIDNCA
HPMYRDYLHRYLENAPGGHITHDLSHVFDLHRNLIAATGSMWG"
complement(5878..6456)
/gene="24258"
/complement(5878..6456)
/gene="24258"
/function="putative regulator; Not classified"
/notes="Residues 1 to 192 of 192 are 99.47 pct identical to
residues 6 to 197 of 303 from Escherichia coli K-12 Strain
MG1655: B2921"
/codon_start=1
/transl_table=11
/product="partial putative transcriptional regulator
LysR-type"
/protein_id="AAG58047.1"
/db_xref="GI:12517453"
/translation="MDIFISKMRNFILLAQTNNIARAEEKIHMTASPFKSTAALEE
QYITFLTRKDNISKAGQYKLFVYQRLSAIDNEIHNSGRSREIVIGDNT
YPTIIFDQILSGDKYEGVTAQVPEFSENGVIDNLDRLDFTIISPHVRSARVOELN
LTISELPLRLNPLVSRVREEQELLOELP"
complement(6623..7363)
/gene="yggE"
/notes="24259"
complement(6623..7363)
/gene="yggE"
/function="phenotype; Not classified"
/notes="Residues 1 to 246 of 246 are 100.00 pct identical
to residues 1 to 246 of 246 from Escherichia coli K-12
Strain MG1655: B2922"
/codon_start=1
/transl_table=11
/product="putative actin"
/protein_id="AAG58048.1"
/db_xref="GI:12517454"
/translation="MKFKVIALAALMGISGMAQAQNELPDGPHIVTSGTASVDVAPDI
ATLAEVNVAAKDAATAKKQADERVAQYISFLELNOIAKKDDISSANLRTQPDYVDG
KSLKGYRAVRVEVTLRLQDLKLSLLDGALKAGLEIRSVSLGVAQPDYKDKARKA
AIDALHQAEQLANGFHRKLGPIVSVRYHVSNTQPSPMVRMKADAPVSAQETTEQA
AIOFDQDVDFVQLEPVDQOAPKTPAAQ"
complement(7456..8091)
/gene="yggA"
/notes="24260"
complement(7456..8091)
/gene="yggA"
/function="orf; Unknown function"
/notes="Residues 1 to 211 of 211 are 98.10 pct identical to
residues 1 to 211 of 211 from Escherichia coli K-12 Strain
MG1655: B2923"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG58049.1"
/db_xref="GI:12517455"
/translation="MFSYVFOGLGALGAAMILPLGPONAFVMNQIRROYHIMALLCA
ISDLVLICGIFGGGALLMQSPWLLALVTWGGVFWLLWYGFCAFKTAMSSNIELASAE
VLKQGRWKIATMLATVNLNPHVYLDTFVVLGGLGQLDVPKPRWFALGTTSASFLWF

```

```

FGLAILAANLAPRLRTAKSORIINLVVCGVMFIALQLARDGIAHAQALFS"
complement(8230..9090)
/gene="yggB"
/notes="24261"
complement(8230..9090)
/gene="yggB"
/function="putative transport; Not classified"
/notes="Residues 1 to 286 of 286 are 100.00 pct identical
to residues 1 to 286 of 286 from Escherichia coli K-12
Strain MG1655: B2924"
/codon_start=1
/transl_table=11
/product="putative transport protein"
/protein_id="AAG58050.1"
/db_xref="GI:12517456"
/translation="MEDLNVDVDSINGAGSWLVANQALLSYAVNIVAALAIIVGLII
ARMISNAVNRMLISRKIDATVDFLSALRYGIIAFTLIAALGRGVQVOTSVIIVLGA
AGLAVGLALQGLSLNLAAGVLLVMPFRPFRAGEVVDLGGVAGTVLSVQIFSTMTMTADG

Query Match      8.7%; Score 62; DB 1: Length 10701;
Best Local Similarity 51.4%; Pred. No. 1.3e-07;
Matches 143; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 49 tcacacctgttgcacgtctcttcgctgcctacgcgcgaacaccatacaggtacgcg 108
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7802 TCAATATTACTATCATCGCTGTTTAAAGCGCCAAACCATACACACAGAGACT 7861

QY 109 atgccaccacgcgcataatcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 168
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7862 ACGCCGCCCGCAGTGACCGCCGACGACGACGACGACGACGACGACGACGACGAC 7921

QY 169 ccacaggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7922 CCAAAATCCCGCGCAATCAGGACCAATCGCTGATCGCGCAAAAGTAGGCAATCAT 7981

QY 229 atagtccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 288
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7982 ATGTGTTACTGACGAGATGTCCTGATTCATCAAAAGCATTTTGTGGCGCGCGGC 8041

QY 289 agtaaaagactgcccccaaaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 326
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8042 AGAATCATAGTGCCTCCCAAGCAAGACGCTTGAAAGTA 8079

RESULT 10
AP002563      266658 bp      DNA      BCT      07-MAR-2001
LOCUS      Escherichia coli O157:H7 DNA, complete genome, section 14/20.
DEFINITION      AP002563 BA000007
ACCESSION      AP002563.1 GI:13363121
VERSION
KEYWORDS
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RMD 0509952)
DNA.
ORGANISM      Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE      1 (sites)
AUTHORS      Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL      20198780
MEDLINE
REFERENCE      2 (sites)
AUTHORS      Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655

```

```

JOURNAL      Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE      20557356
REFERENCE    3 (sites)
AUTHORS      Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
              Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
              Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
              Shinagawa,H.

TITLE        Complete nucleotide sequence of the prophage Vri-Sakai carrying the
              O157:H7 strain derived from the enterohemorrhagic Escherichia coli
              Gene 258 (1-2), 127-139 (2000)

JOURNAL      4 (sites)
MEDLINE      20564182
REFERENCE    4 (sites)
AUTHORS      Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
              Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
              Tanaka,M., Tohe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
              Ogawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
              Shinagawa,H.

TITLE        Complete genome sequence of enterohemorrhagic Escherichia coli
              O157:H7 and genomic comparison with a laboratory strain K-12
              DNA Res. 8 (1), 11-22 (2001)
              21156231

JOURNAL      5 (bases 1 to 266658)
MEDLINE      Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
REFERENCE    Hayashi,T.
AUTHORS      Direct Submission
              Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
              Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
              Japan [E-mail:ken@gen-info.osaka-u.ac.jp,
              URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
              Fax:81-6-6879-2047]
              genome project.

COMMENT      Location/Qualifiers
FEATURES     1..266658
              /organism="Escherichia coli O157:H7"
              /strain="O157:H7"
              /sub_strain="RIMD 0509952"
              /db_xref="taxon:83334"
              complement(244..693)
              /gene="ECS3650"
              complement(244..693)
              /gene="ECS3650"
              complement(244..693)
              /note="similar to B2790_ECOLI gi|1789154 percent identity
              100 in 149 aa (Conserved in E.coli K-12)"
              /codon_start=1
              /transl_table=11
              /evidence=not_experimental
              /product="hypothetical protein"
              /protein_id="BAB37073.1"
              /db_xref="GI:13363122"
              /translation="MAEIGIFVTGMYGNSLVAEAEAILTAQGHKATVFEDPELSDW
              LPYQKYLIVVTSTTGGDLPSIVPLFGIKDSIGFQPNLRYGVIALGDSYVNFNC
              GKGQFDALQEQSAQRVGEMLLIDASENPETESNPWEQWGTLTS"
              complement(711..1493)
              /gene="ECS3651"
              complement(711..1493)
              /gene="ECS3651"
              /note="similar to YQCD_ECOLI gi|1789155 percent identity
              99 in 260 aa (Conserved in E.coli K-12)"
              /codon_start=1
              /transl_table=11
              /evidence=not_experimental
              /product="hypothetical protein"
              /protein_id="BAB37074.1"
              /db_xref="GI:13363123"
              /translation="MLEILYODEWLVAVNKPQSGWLVRHSWLRDRKVVVMOTVRDQIG
              OHVFAHRLDRPTSGVLLMGLSSSEAGRLLAQFQEQHQRKRYHAIVRWLWAEVLDY
              PLVEEDKADKFAREDGQPVATRYGLATFEMPVATGRYPIYGLVELEPKTR
              HQHLRHLLAHRPIIGDSKGDLRQNSGAHFGLQRLMLHASLSLTHPTGEPFL
              IHQLDDTWQALSQFGWRGLLPENERVEFSAPSQDGERSS"
              complement(1493..1822)
              /gene="ECS3652"
              complement(1493..1822)

JOURNAL      105-117i-1_copy_1016_1726.rge
MEDLINE
REFERENCE
AUTHORS

TITLE

JOURNAL      98 in 109 aa (Conserved in E.coli K-12)"
MEDLINE      /gene="ECS3652"
REFERENCE    /note="similar to B2792_ECOLI gi|1789156 percent identity
              98 in 109 aa (Conserved in E.coli K-12)"
              /codon_start=1
              /transl_table=11
              /evidence=not_experimental
              /product="hypothetical protein"
              /protein_id="BAB37075.1"
              /db_xref="GI:13363124"
              /translation="WTHDHRVLOLOALEALLREHOHWNRDEPQHPQFNSTQPFEMDT
              MEPLLEQWLVIPRMHDLNNNQPLPGFAVAPYIEMALATDHFQALILAELEKLDA
              LFADDAS"
              complement(1865..2233)
              /gene="ECS3653"
              /note="inhibitor of CsrA"
              /product="CsrB RNA"
              complement(1865..2233)
              /gene="ECS3653"
              complement(2444..2989)
              /gene="ECS3653"
              complement(2444..2989)
              /note="similar to SYD_ECOLI gi|1789157 percent identity 98
              in 181 aa (Conserved in E.coli K-12)"
              /codon_start=1
              /transl_table=11
              /evidence=not_experimental
              /product="Syd protein"
              /protein_id="BAB37076.1"
              /db_xref="GI:13363125"
              /translation="MDDLTAQALKDFTARYCDANHEEHKSWPLSEELYGVSPCIIST
              TEDAVYQWPQFTGEQNVNAVERAFDIVQTIHTFTYTFAGDMHAFQGIKLTLLQ
              TWSEDDFRVQENLGHVLTQKRLKPLTPTLTATLEELVISCNLSGEVCKETLGT
              RKRTHLAPNLAELNLQKPLI"
              3057..3905
              /gene="ECS3654"
              3057..3905
              /gene="ECS3654"
              /note="similar to YQCD_ECOLI gi|1789158 percent identity
              98 in 282 aa (Conserved in E.coli K-12)"
              /codon_start=1
              /transl_table=11
              /evidence=not_experimental
              /product="hypothetical protein"
              /protein_id="BAB37077.1"
              /db_xref="GI:13363126"
              /translation="MSSYVANOALAGLTGKSTDYRDTYDASLLQGVPRSLNRPDLGL
              KADNLPQGTDIWTLVELSWLNAGLQVAVGHVELDYTSVNLIESKSKLYLNSFNO
              TRFNWDEVROTLERLDSTCAQGEVSVALYRLDELEGOPIGHNGTCIDDDQDIIDNY
              EFTTDYLDENATSGERKVEETLVSLKSNCLITHQPDWGSIOIQIRGQIDREKLLRY
              LVSPFRHNEFHEQCVERIFNDLLRFQCEKLSVARYTRRGGLDINPWRNSDFVPST
              TRLVRQ"
              4017..5381
              /gene="ECS3655"
              4017..5381
              /gene="ECS3655"
              /note="similar to YGDH_ECOLI gi|1789159 percent identity
              99 in 454 aa (Conserved in E.coli K-12)"
              /codon_start=1
              /transl_table=11
              /evidence=not_experimental
              /product="hypothetical protein"
              /protein_id="BAB37078.1"
              /db_xref="GI:13363127"
              /translation="WTHISPLGSMDSLSQLEVDMLKRTASSDLYQLFRNGSLAVLNS
              GSLTNSKELLSRFENFDINVLRRERGKLEINPPPEAFVDGRIIRALQANLFAVL
              DILFYGGIHNLTVPFNPDLNLSNVEHITNLVEILNARALHVGEAPNVCWGHSIN
              ENEDYARVGNQGLRELINICTGCGPGAMEAPKGAHVGAHQRYKDSRFIGTEPS
              ITAAAPPPLVNELIMDIEKRLAEFVRIAGHIIIFPGVGTAEEELLLGLMNP
              NKDQVPLILTPKESADYFRVLDYFVHTLGENARRHYRIIIDAAEVAROKMKMP
              LVKENRDTGDAYSFNWSMRTAPDQMPFEPESHENMANLKLYDQPPVLAADLRRAF
              SGIVAGNVKEVGIRAEIEFGPYKINGDKKEIMRRMDLLQGFVAQHRMKLPGSAYIPCY
              EIC"

```

```

gene      5938..7227
/gene="ECs3656"
CDS      5938..7227
/gene="ECs3656"
/note="similar to SDAC_ECOLI gi|1789160 percent identity
100 in 429 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="probable serine transporter"
/protein_id="BAB37079.1"
/db_xref="GI:13363128"
/translation="METQSTIASKDSRSNRKTDWMLGLYTAIGAGVLELPIN
AGVGMPLIALLAPMTFFAHRGTRFVLSKNGEDITEVEEHFGIGAGKLIT
LLYFAIPLLYSVAITNTVFMSHQLMGTPPRAILSLIIVGMGTIVRGEQM
IVKAMSLVFFVGLMLLALYLPQWNGALETSLDTSATSGLWMLTILAIQPM
VFSFNHSPIITSSFAVAKREYDMAEQCKILAFAHIMVLTVMFVFVSCVLSLTPA
DLAAKQNTSILSYLANHFNAPVIAWMAPIATIAITKSLFGLHLCARGFNGMVIK
SLRKGKGSIEINKNLTALEMLVTTWIVATINPSILGMETLGGPIIIMLFLMPHY
AIQKVPAMRKYSIGHISNVVVVVMGLIAISAFISLFS"
7285..8652
/gene="ECs3657"
7285..8652
/gene="ECs3657"
/note="similar to SDAA_ECOLI gi|1788116 percent identity
77 in 455 aa, similar to SDAB_ECOLI gi|1789161 percent
identity 99 in 455 aa (Conserved in E.coli K-12)"

Query Match      8.78; Score 62; DB 2; Length 266658;
Best Local Similarity 51.48; Pred. No. 2.6e-07;
Matches 143; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy 49 tccacctgttgcctgctcttgcctgacgacgcaaacacacaggttaagcg 108
Db 161304 TCATATTACTACTCATCGCTGTTTAAAGCGCCAAACCATACCAGCAAGAGACT 161363

Qy 109 atgcaccccgccgataatcagacgacgacgacgacgacgacgacgacgacg 168
Db 161364 ACCGCGCCCGAGGTACACGCGCCGACCAACACGCGGACTCATCAATACGCGGTACCA 161423

Qy 169 cccaaagtgccgagatgaacaaagacgacgacgacgacgacgacgacgacgacg 228
Db 161424 CCAAAATCCCGCGCAATCAGGACCAATTCGCTGATAGCGCAAGTAAGGCAATCATATA 161483

Qy 229 atgagtccttcgcttaattccttctgttataccagctacatctgcgtccgacgac 288
Db 161484 ATGTGGTACTGACGACGAATGCCCTGATTCATCACAAAGCATTTGTGGCGCGCGGC 161543

Qy 289 agtaaaagactgcccccaagacgacacgtgtaataatgaa 326
Db 161544 AGAATCATAGCTGCCCAAGACGAAGACCTTGAAAGTA 161581

RESULT 11
LOCUS      MTCY39      38500 bp      DNA      BCT      17-JUN-1998
DEFINITION      Mycobacterium tuberculosis H37Rv complete genome; segment 89/162.
ACCESSION      Z74025 AL123456
VERSION      Z74025.1 GI:3261586
KEYWORDS
SOURCE      Mycobacterium tuberculosis.
ORGANISM      Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 38500)
Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.B.,
Tekaita,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
Hamlin,N., Holtroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skellton,S.,
Squares,S., Sqaures,R., Sulston,J.E., Taylor,K., Whitehead,S. and
Barrell,B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)
98295987
Erratum: [[published erratum appears in Nature 1998 Nov
12:396(6707):190]]
2 (bases 1 to 38500)
Parkhill,J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:1403436.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a hidden Markov Model of TB genes
implemented in TParse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
FEATURES
source      Location/Qualifiers
1..38500
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
<1..38419
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y39"
/complement(91..777)
/gene="mpt64"
/gene="mpt64"
/note="RV1980c, (MTCY39.39), len: 228. mpt64 identical to
MPB6_MYCO P19996 immunogenic protein mpb64/mpt64 (228
aa). Similar to Rv3036c, (MTV012.51c)"
/codon_start=1
/transl_table=11
/product="mpt64"
/protein_id="CAA98382.1"
/db_xref="GI:1403475"
/db_xref="SWISS-PROT:P19996"
/translation="MRKIFMLVAVVLLCCSGVATAAPKTYCEELKGTDTGACQIQ
MSPAYNINISLPSYYPDOKSLNLTQATDKFLSAATSTPREAPYELNITSATQOS
AIPRGTOAVLVAYONAGTHTTYYKAFDWDQAYRKPTTYDTLQADTDLPLVPVFP
IVQGLSKQTGQQVSTAPNAGLDPVNYQNPVAVTNDGVIFEFNGELLPEAAGTQVLY
PRSAIDSMLA"
complement(787..791)
/note="ggagg, potential rbs upstream of Rv1980c"
complement(968..1936)
/gene="nrdF"
complement(968..1936)
/gene="nrdF"
complement(968..1936)
/gene="nrdF"
/note="RV1981c, (MTCY39.38), ribonucleotide reductase
small subunit, len: 322, highly similar to RIR4_SALTY
P17424, ribonucleoside-diphosphate reductase (319 aa)
FASTA scores, opt: 1402, z-score: 1716.0, E(): 0, (66.08
identity in 315 aa overlap); Also similar to
Rv3048c, (MTV012.63c); containsPS00368 Ribonucleotide

```

reductase small subunit signature"

```
/codon_start=1
/transl_table=11
/product="nrdf"
/protein_id="CAA98381.1"
/db_xref="GI:1403474"
/db_xref="SWISS-PROT:Q10840"
/translation="MTGKLVVERHAINNRLLDAKDLQWERTLGNFWLPKIPLSND
LASWOTLSTEQTTIRVFTGLTDLDTAQATVAVAMIDDVTPHEEAFLTNMAFME
VHAKYSSTFSLCTCDDAFDSEQNQLQKRAQIIVDYRGDDALKRKAQSSVML
ESFLYSGFYLPMYSSKGLNTADLRLIIRDRDAVHGVIYGYKCORGLADLTDAER
ADHRYETCELLTLYLANEIDYAHDLDELGTDDVLPYMYRNANKALNLGQPAFDR
DTCQNPVRAALDPAEAGHDFSSGSSSYVMGTHQPTTDDWDF"
/complement(1589..1636)
/gene="nrdf"
/feature="misc_feature"
```

misc\_feature

/note="PS00368 Ribonucleotide reductase small subunit signature"

```
/complement(1946..1949)
/note="aggga, potential rbs upstream of Rv1981c"
/complement(2161..2580)
/gene="Rv1982c"
/complement(2161..2580)
/gene="Rv1982c"
```

/note="Rv1982c" (MTCY39.37), len: 139. Similar to Mycobacterium tuberculosis 292772|MY20H10\_5 (131 aa) FASTA scores: opt: 288 z-score: 388.5 E(): 4.1e-14; 40.2% identity in 127 aa overlap, also similar to Rv2759c, and Rv0609"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein Rv1982c"

/protein\_id="CAA98401.1"

/db\_xref="GI:1403473"

/db\_xref="SWISS-PROT:Q10874"

/translation="MIVTSVAVLVQGERPHATVAAALAGAHSPVMSAPTVAECLI

VLTAHQVAPRIFFELRSEIGLSVSTAEHAATQRAFLRYGKGRHRAALNFGDCM

TYATQHQHPLLVANGNPPQTDLEFRGVVGPVGA"

2981..2985

/note="ggagj, potential rbs upstream of Rv1983"

2992..4668

/gene="PE\_PGSR"

2992..4668

/gene="PE\_PGSR"

/note="Rv1983" (MTCY39.36c), len: 558; Member of M.

tuberculosis PE\_PGSR family. Contains PS00141 Eukaryotic

and viral aspartyl proteases active site; similar to

MTCY130.10c, (43.3% identity in 522 aa overlap)"

/codon\_start=1

/transl\_table=11

/product="PE\_PGSR"

/protein\_id="CAA98400.1"

/db\_xref="GI:1403472"

/db\_xref="SWISS-PROT:Q10873"

/translation="MSFLVVPFLTSAADVENIGSLRANAAASTTALAAGA

DRVSAALPAFRCQVQASAFHQFQTLNSAGSFAAEATIASQLTAQ

HDLLGAVALPTELLRGLDGPATATSPNGAGGLLYNGNGYSATASGVGGA

GGAGLNGAGGAGGAGNAPGGAGGGLNGGIGGPGGASSIPMGSGAGGTGG

CUTI\_ALTBFR P41744 cutinase precursor, (209 aa) FASTA scores: opt: 283, z-score: 322.0, E(): 2.2e-11, (32.6% identity in193 aa overlap). Also similar to M. tuberculosis Rv3452, Rv3451, Rv2301, Rv1758, Rv3724."

```
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1984c"
/protein_id="CAA98399.1"
/db_xref="GI:1403471"
/db_xref="SWISS-PROT:Q10837"
/translation="MTPSLRVIVGVVATTALVSPAGGRAAHADPCSDIAVVFAR
GTHASGLGDGEAFVDSLTQVGRSGICVYAVNVPASDDYRASNSDDASAHQIR
TVASCPNTRIVLGVSQATIDLTSTAMPAVADHVAALFPGSPGSSMLWGGG
SLPTGLPYSKTINLCAPDDPITCGGNIMAHVSIVQSGMTSQATTAANRLDHAG"
/complement(4950..4967)
/gene="Rv1984c"
/feature="misc_feature"
```

/note="PS00155 Cutinase, serine active site" complement(5739..6650) /gene="Rv1985c" complement(5739..6650) /gene="Rv1985c"

/note="Rv1985c" (MTCY39.34), len: 303, LysR family member similar to many regulatory proteins, especially ICIA\_ECOLI P24194, chromosome initiation inhibitor, (297 aa), and P94632 LYSINE EXPORT REGULATOR PROTEIN. (290 aa). FASTA scores, P94632 LYSINE EXPORT REGULATOR PROTEIN opt: 705 z-score: 850.6 E(): 0; 42.7% identity in 288 aa overlap; and ICIA\_ECOLI P24194 opt: 520, z-score: 632.5, E(): 1.1e-28, (35.8% identity in 285 aa overlap); contains PS00044 Bacterial regulatory proteins, LysR family signature. Also contains helix-turn-helix motif at aa 22-43, (+5.52 SD)."

/codon\_start=1

/transl\_table=11

/product="hypothetical protein Rv1985c"

/protein\_id="CAA98410.1"

Query Match 8.7%; Score 61.6; DB 3; Length 38500; Best Local Similarity 55.0%; Pred. No. 2.3e-07; Matches 165; Conservative 0; Mismatches 129; Indels 6; Gaps 2;

QY 418 caactggacacatttctctcaatcattgatgaagcagcttcgaagggccctccttagcc 477

Db 6623 CAGCTGCGCGCATTTGCTCCGCTGCTGCACTGGCGAGCTTCGATCGCGCCGGGAGCGC 6564

QY 478 ctttcatttccccctcgccggtgagtcagcgcgttaaagctctcagcatcacgtgggt 537

Db 6563 CTACATGTCACCCCGTGGCTGCTCAGTCAGCGCATCAAGTCGTTGCGAGCAGGTTCGCG 6504

QY 538 cgagtgtgtatcgcgaccccaaccgcccgaagcaaccgaagcggtgaagtccttctg 597

Db 6503 CAGGTGCTGCTGCTCAGGAAAAAGCCATCTCGGGCGACGACGAGTATCCGCTGTTG 6444

QY 598 caagcagcgcggaataatgtgtgtgtgcaagcagaaactaaagcgcaactatctgg--ac 655

Db 6443 CGTTGGCGCGCAACACACGCTGCTCGAGTCCGAGCGCTCGCTGAAATGGTGGCAAC 6384

QY 656 gcttgctgtaa----atcccgtaaacatcgccatcaacgagattcgatcccatgg 711

Db 6383 GCGTCTGCTGAAACGACGCGGATCCATTCGTTAAGCCGATTCATGCGGACATGG 6324

misc\_feature

/gene="PE\_PGSR"

/note="PS00141 Eukaryotic and viral aspartyl proteases active site"

complement(4656..5309)

/gene="Rv1984c"

complement(4656..5309)

/gene="Rv1984c"

/note="Rv1984c" (MTCY39.35), len: 217. Probable cutinase precursor with N-terminal signal sequence, similarity to

RESULT 12

AX030085/c

LOCUS

AX030085

DEFINITION

Sequence 15 from Patent EP1016710.

AX030085

ACCESSION

AX030085.1

GI:10190302

AX030085

636 bp

DNA

PAT

16-SEP-2000

Sequence 15 from Patent EP1016710.

AX030085

AX030085.1

GI:10190302

Escherichia coli.

Escherichia coli

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Escherichia.

```

REFERENCE 1 (bases 1 to 636)
AUTHORS Nakanishi,K., Aleshin,V.V., Livshits,V.A., Tokhmakova,I.L.,
Troshin,P.V. and Zakataeva,N.P.
TITLE Method for producing L-amino acids
JOURNAL Patent: EP 1016710-A 15 05-JUL-2000;
AJINOMOTO KK (JP)
FEATURES
    source      Location/Qualifiers
        1..636
            /organism="Escherichia coli"
            /db_xref="taxon:562"
        1..636
            /note="unnamed protein product"
            /codon_start=1
            /transl_table=11
            /protein_id="CAC09317.1"
            /db_xref="GI:10190303"
            /translation="MFSYFQGLALGRAAILPLGPONAFVNMNOGIRROYHIMTALLCA
ISDYVLCAGIFGSGALLMQSPWLLALVTWGGVAFLLWYGFYGFAPKMTAMSNIELASAE
VMKGRWKKIATMLAVTLPNPHVYLDTFVVLGSGGOLDVPEPKRWFGALGTISAFSLWF
FGALLAAWLAPLRTAKAQRINLVVGCVMWFIALQLRDGTAAHAQALFS"
BASE COUNT 120 a 143 c 180 g 193 t
ORIGIN

Query Match      8.5%; Score 60.4; DB 9; Length 636;
Best Local Similarity 51.1%; Pred. No. 2.2e-07;
Matches 142; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 49 tccacctgttgcatgcgtcttctgcctgcatgcagcgcaaacaccataacagtgtaacg 108
Db 230 TCATATTACTGCTCTATTCGCTGTTTAAAGCGCCAAACCATCACAGCAAGAAGCGT 231
Qy 109 atgcacccccagcgcataatatcagcagcagcgcgcgcgcattggacaaaagatcaacg 168
Db 230 ACGCGCCCCAGGTGACCCAGCGCCAGCAACACGCGCGACTGCATCAATACGCGTGCCA 171
Qy 169 cccaaggtgcgcggatgaacaaaagacgctcagaataataacacagagagaacacgca 228
Db 170 CCAAAAATCCCGCGCAATCAGGACCAAAATCGCTGATAGACACAAAGTAAGCGCAATCATA 111
Qy 229 atgagtccttcgcgttaattccttctgttataccagcatagatcgtcgctccgatgac 288
Db 110 ATGTGGTACTGACGACGTATCCCTGATTCATCATCAAAAGCATTTTGTGGACCGCGGT 51
Qy 289 agtaaaagactggcccccaaaagcagacctgttaatgaa 326
Db 50 AGGATCATAGCGGCCCCCAAGTGCAAGACCTGTGAAAGTA 13

RESULT 13
ECFAPGK/c
LOCUS      ECFAPGK      8029 bp      DNA      BCT      21-MAR-1994
DEFINITION Escherichia coli fda, pgk and gapB genes for fructose
1,6-bisphosphate aldolase (class II), phosphoglycerate kinase and
glyceraldehyde 3-phosphate dehydrogenase.
ACCESSION X14436.1 GI:41417
VERSION   X14436.1 X14682
KEYWORDS  fda gene; fructose 1,6 biphosphate aldolase; gapB gene;
glyceraldehyde 3-phosphate dehydrogenase; pgk gene;
phosphoglycerate kinase; unidentified reading frame.
SOURCE    Escherichia coli.
ORGANISM  Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE 1 (bases 1 to 8029)
AUTHORS Alefounder,P.R.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-1989) Alefounder P.R., University of Cambridge,
Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, England
2 (bases 4376 to 5508)
REFERENCE Alefounder,P.R., Baldwin,S.A., Perham,R.N. and Short,N.J.
AUTHORS Cloning, sequence analysis and over-expression of the gene for the
TITLE class II fructose 1,6-bisphosphate aldolase of Escherichia coli

JOURNAL Biochem. J. 257 (2), 529-534 (1989)
MEDLINE 89193446
COMMENT Data kindly reviewed (28-JUN-1989) by Alefounder P. R.
FEATURES
    source      Location/Qualifiers
        1..8029
            /organism="Escherichia coli"
            /strain="CS520"
            /db_xref="taxon:562"
            /map="62 min"
            /clone="pLC33-5"
        142..429
            /note="ORF1 (AA 1-95)"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA32600.1"
            /db_xref="GI:41418"
            /db_xref="SWISS-PROT:P21437"
            /translation="MQAELIDFCQAKGDYTENRQIAEQEKRCAMGVQDVNRYSLDE
LVRGNDILFATGVTGGELVNGIQTANGVRTQTLLGGADQTCNIIDSLH"
        172..183
            /note="inverted repeat A"
        292..303
            /note="inverted repeat A"
        517..918
            /note="ORF2 (AA 1-133)"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA32601.1"
            /db_xref="GI:41419"
            /db_xref="SWISS-PROT:P11663"
            /translation="MKTATHILLQGIHQFLPSLFDVNDSEIVEYAKPKLLAQSGPLDD
IDVALRLIYALGRMDKWLVIADITHFSQFWHLNEQDETGFADMTWDFISNVNSITR
NAMLYDALKAMKFAFWSVSEARFSGMKVTA"
        648..657
            /note="inverted repeat B"
        680..690
            /note="inverted repeat C"
        718..729
            /note="inverted repeat C"
        800..811
            /note="inverted repeat B"
        1196..1669
            /note="ORF3 (AA 1-157)"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA32602.1"
            /db_xref="GI:41420"
            /db_xref="SWISS-PROT:P11664"
            /translation="MDGFHHNSWLDHQLRPFKGAPEETDVAKLNTENRQVVEGDC
WPQYDRKHDPVEDALHVTAPLVIVEGNWLLDDEKWLKELASCFDSIFIHAPAQLUR
ERLTSRKIAGGLTRQVAEAFYARTDGNVSRVLMNSQANLIVEMTEEGRYHFTS"
        1741..1761
            /note="putative catabolite gene activator protein binding
site"
        1881..1906
            /note="inverted repeat D"
        1915..1942
            /note="inverted repeat D"
        1954..2973
            /note="glyceraldehyde 3-phosphate dehydrogenase (AA
1-339)"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA32603.1"
            /db_xref="GI:41421"
            /db_xref="SWISS-PROT:P11603"
            /translation="MTVRVAINGFRGNRVVALYESGRRAETIVVAIINELADAAGM
AHLKDYTSHGFAWEVRQDQLFVGDDAIRVHRSLSQSLPWRGLGVQVVDCTGY
VSGREGEAHAAAGCAKKVLFSPGSGNDLATVYGVNQDLRRAHSRIVSNACSTNCI
IPVILKLDIAYGIESGTVTIHSAHMDQQQVIDAYHPDLRRAASOSIIPVDTKLAAG
ITRFPPQNDRFEAIAVRVPTINVTAIDLSTVKPKVKAENEVLLQLKAAQAFHGIV
DYTELPLSVDFDNHDPHSAIVDCTQTRVSGAHLIKTLKILVWCDNEWGNRMLDPTLAA
TVAFR"

```

```
CDS
3023. .4186
/codon_start=1
/transl_table=11
/protein_id="CAA32604.1"
/db_xref="GI:41422"
/db_xref="SWISS-PROT:P11665"
/translacion="MSVITMDLDLAGKRVFIRADLNVPVKGKVTSDARISLPTI
ELALQKQAVNMTKDETELSKKAALCDVFMDAEGTAHRAQAASHGIGKPADV
AGELVLENNVSKGKDETELSKKAALCDVFMDAEGTAHRAQAASHGIGKPADV
ACAGPLLAELDALGKALKEPARPVAIVGGSKVTKLVLDLSKIAQLIVGGIA
VFIAQGLDIDGSKSYEADLVDEAKRLTTCNIPVSDVRVATEFSETAPATLKSND
NFKADQLIDGASQAEALAIKNAKTILWNGPVGVEFPNFKETETIVANAIDSEA
FTAGGGDTLAAIDLFIADKISYISTGGCAFLEFVEGKVLPAVAMLEERAKK"
4401. .5480
/note="fructose 1,6-biphosphate aldolase (AA 1-359)"
/codon_start=1
/transl_table=11
/protein_id="CAA32605.1"
/db_xref="GI:41423"
/db_xref="SWISS-PROT:P11604"
/translacion="MSKIFDVKPGVITGDDVQKVOVAKENNEALPAVNCVGTDSIN
AVLETAARKVAVIVQFNSNGGASFTAGKGVKSDVFGGAALGAIAGAHVHOMAEHYG
VPYLHTDCAKLLPWIDGLDAGEKHFATGKPLFSSHMIDLSESLQENIEICSK
YLERMSKICMTLEIELGTCGDEEDVDNSHMDASALYTQPEDVDYATYELSKISPRF
IAASGCVNHYKPGNVLTPTILRDSQEVYVSKHNLPHNSLNFVFGHGGSGTAQEI
DSVSYGVNKNIDTDTQWATWEGVLNYKANEAYLQGLGNPKGEDQPNKKYYDPRVW
LRAGQTSMTARLEKAFQELNAIDL"
5545. .5573
/note="repetitive extragenic palindromic sequence"
5580. .5586
/note="inverted repeat E"
5597. .5603
/note="inverted repeat E"
complement(5613..5762)
/note="repetitive extragenic palindromic sequence"
5838. .6698
/note="ORF4 (AA 1-197)"
/codon_start=1
/transl_table=11
/protein_id="CAA32606.1"
/db_xref="GI:41424"
/db_xref="SWISS-PROT:P11666"
/translacion="MEDLNVDPSINGAGSLWVANQALLSYAVNVAIAIIVGLII
ARMISNAVRLMSKDATVAFSLVRYGIIAFTLTAALGRVGVOTASVIAVLGA
AGLVGLALQSLNSLAAGVLAIVMRPEPAGVYDLGGVAGTVLSVQIFSTPMRADG
KLIIVPNGKIIAGNIINFSRPNREFTIIGVAYSDIDQVAKOIFNLTKIISDRILKD
REMTVRLNELGASSINFFVVRVWNSGDLQNVYMDVLRIKREFDAAGISFPYPQMDVN
FKRVKEDKAA"
6879. .7472
/note="ORF5 (AA 1-197)"
/codon_start=1
/transl_table=11
/protein_id="CAA32607.1"
/db_xref="GI:41425"
/db_xref="SWISS-PROT:P11667"
/translacion="MILPGQNAFVNGQIRROVHIMIALCAISDLVICAGIFGG
SALLMQSPMLIALVTGGVAFLLWYFGFAKTMSSNIFLAESAAYMKGRWKIIATML
AYTLNPHVYLDFTFVVLGSLGQLDVEPKRWFALGTISASFLWFFGLALLAALAPRL
FPAKQRIINLVGCVGMFIALQLARDGIAHAQALFS"
7565. .78029
/note="ORF6 (AA 1-155)"
/codon_start=1
/transl_table=11
/protein_id="CAA32608.1"
/db_xref="GI:581079"
/db_xref="SWISS-PROT:P11668"
/translacion="MKFKVIALAALMGISGMAQAQNELPDGHVIVTSGTASVDVAPDI
ATFLAEVNAVNAKQADRVAYQISFLELNQIAKRIAKISSANLRTQPDYQDQG
KSLIKGYRAVRVTEVTLRQLDKLSLDDGALKAGLNEIRSVSLGVAQPDAYKD"
1984 a 1922 c 2114 g 2009 t
BASE COUNT
ORIGIN
```

```
Query Match 8.5%; Score 60.4; DB 2; Length 8029;
Best Local Similarity 51.1%; Pred. No. 3 8e-07;
Matches 142; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 49 tccacettgttgcctgctgttcctgcatgacggaacaccataacaggttaagcg 108
Db 7126 TCAATATTACTGCTCATTTGTTTAAAGGCGCAAAACCATACACAGCAAGGCT 7067

QY 109 atgccacccacgcataataatagacagcatcgcgcgccattggacaaaagatcaag 158
Db 7066 ACCGCCCCCGAGGTGACAGCGCCACACACGCGGACTGCATCAATACGCGCTGCCA 7007

QY 169 cccaagggtcgccgatgaacaaaaagagcgtcagaattaaacacacgagagaacccga 228
Db 7006 CCAAAATCCGGCGCAATCAGGACCAATCGCTATAGCACAATAAGGAATCATATA 6947

QY 229 atgagtccttcgcgttaattcctgtttaaaccacagtcattcgcggtcgatggac 288
Db 6946 ATGTGCTACTGACGACGTATGCTGCTGATTCATCATAAAGCATTTTGTGGACGCGGT 6887

QY 289 agtaaaagactgcccccaaaaagacagacacctgtaataaa 326
Db 6886 AGGATCATAGCGCCCAAGAGTGCAGAGACCTTGAAGATA 6849

RESULT 14
LOCUS AE000375 10362 bp DNA BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 265 of 400 of the complete
genome.
ACCESSION AE000375 U00096
VERSION AE000375.1 GI:1789282
KEYWORDS Escherichia coli K12.
SOURCE Escherichia coli K12.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 10362)
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Colliado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goaden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 10362)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE 3 (bases 1 to 10362)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE 4 (bases 1 to 10362)
AUTHORS Plunkett, G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
```

```

/translation="MINATLAESTIRRLROGERATLAQAMTILVESRHPHQAALSTQLL
DAIPYQNTLRVLGVTGPGAGKSTFLFAFGMLLRIGELIKVAVIADVPSSPVTVGGSI
DKYTRMDLARAEEAFIRPSPSGHLGASQARMLMCEAAGYDVVIVETVGVGQS
ETEVARWDCSTISYIAGCGDQLQIKKGLMEVAIDILNKDDGNHTNVALTARMYE
SALTILRRKYDEWOPRVLVTCSALEKRGIDELWAIDILDKTALTASGRQLQVQOQSVE
WLKQTEEVLNHLFANEFDRIYRQTLVAVKNNTLSPRTGLRQLSEFIQTQYFD"
4299..5126
/genes="ygfG"
/feature="putative enzyme: Not classified"
/note="o275; this 275 aa ORF is 30 pct identical (6 gaps)
to 250 residues of an approx. 296 aa protein ECHM_RAT_SW:
P14504"
/codon_start=1
/transl_table=11
/product="putative enzyme"
/protein_id="AAC75956.1"
/db_xref="GI:1789286"
/translation="MNSURNRNLIKIPMSYQYVNVTVINKVAVIEFNYGKRKNALSK
VFIDLMQALSDLRNATRCIRILRAPSGSKYFSGAGDHIELPSSGRDPLSYDDPLRQI
TRMTQKFPKPIISVMEVGSGAGFAMMSDILLIAASTSTSMTPNVLGSPVYNLVGIIH
NLTRDQAGHIVKELIFTASPTAORALAVGLNHVVEVELEDFTLQMAHHITSEKAPL
AIAYIKELRYLGEAHTWNSDEFERIQCMRRRAVYDSEDYQECMNAFLEKRPKFVGH"
5150..6628
/genes="ygfH"
/note="o492; this 492 aa ORF is 46 pct identical (7 gaps)
to 490 residues of an approx. 544 aa protein CATL_CLOK1
SW: P349946"
/function="putative enzyme: Not classified"
/note="o493; this 492 aa ORF is 46 pct identical (7 gaps)
to 490 residues of an approx. 544 aa protein CATL_CLOK1
SW: P349946"
/codon_start=1
/transl_table=11
/product="putative coenzyme A transferase"
/protein_id="AAC75957.1"
/db_xref="GI:1789287"
/translation="METQWTRMTANEAAEIIQNDMDVAFSGFTPTGSPKALPTAIRR
ANECEHAKPKYRIELLTGCASISAAADDVLSADAVSWRAPYQTSGLGRKKINQGVASF
VDLHSEVAQWNVYFGGFDIDVAVTEASALPDGRVMT.TSGIGNAPTLLLRKAKVIEE
LNHYHDRVAELADIVIPGAPPRNSVSIFFHAMDRVCTRYQVIDPKKLVAVYETNLPD
AGNHDDCPQMCQIADNVPTFLIQEMAHGRIPPEFLPQSGVGNINNAVMALGENP
VIPPFWMYSEVLYSVALHETGKISGASASLTSIADSLRKIDYNDMDYFASRIVLRP
SITSNNEPIIRRLGVNLVLEGTDIYGHANSTHVAGDVLMMNGIKSGSGDFERNAYLSI
FMPSYIAKEGTSITVPMCKEAGHSEHSVKVITIEQGLADLRGLSPLOQRARTIIDNCA
HPMYRDYLHRYLENAPEGHIIHDLSHVEDLHRLNLATGCSMLG"
complement(6625..7536)
/genes="ygfI"
/note="b2921"
complement(6625..7536)
/genes="ygfI"

```





```
misc_feature
GETAIVQIRNREMPVKVTKPVFVRNGKAVA"
/complement(5370..10317)
/notes="similar to GenBank Accession Number D90281
(ECOPUV)"
gene
complement(5506..6708)
/gene="visc"
CDS
complement(5506..6708)
/gene="visc"
/notes="ORF_f400"
/codon_start=1
/transl_table=11
/protein_id="AAAG69074.1"
/db_xref="GI:882436"
/translation="MOSVDVAIVGGMVLAVAGCGGSLRVAVLEQVQVEPLAANA
PPQLRVSAINAAEKLTLRLGVWQDILRRASCYHGMVWMDKDSFGHISFDDQSMGYS
HLGHIVENSVIHYALNKAHQSSDITLLAPAEIQQVAMGENETFLTLDGCSMLTARLV
IGADGANSWLRNADIPLTFWDYOHVALVATIRTEPHDAVARQVFHGEIGLAFPLPS
DPLHCSIVWSLSPPEAQMOQASEDEFNRALNIAFONRLGLCKVESAROVFPLTGYVA
RPFASHELALVGDAHTIHPLAGQGVNLGPMDAELIAELKRLHROCKDQGVLYLRR
YERSKRHSALMAGMQGFRDLFSGTNPARKLLRDIGKLADTLPGVKPOLIRQAMGL
NDLPFWLR"
/complement(6731..7909)
/gene="ubih"
gene
complement(6731..7909)
CDS
/gene="ubih"
/notes="ORF_f392"
/codon_start=1
/transl_table=11
/protein_id="AAAG69075.1"
/db_xref="GI:882437"
/translation="MSVIVGGMGATLALISRLSHGALPVHLIEATAPESHAPHG
FDGRALAAAGTCQQLARFQWGLADCATITTVHSDRGHAGFVTLAAEDYOLAAAL
GOVEHLNVGQFALLRKAPVTLHCPDRVANVARTQSHVEVTLSEGETLGRVLVA
ADGTHSALATACQVDWQOEPEQLAVIANVATVAHEGRAFERFTQHGPLMLPMSDG
RCSLVCHPEREREVLSWSEDFKRELOSAFGWRLGKITHAGKRAYPLALTHAARS
IHTWLVGNAAOTLHPHAGQGNLGMWVMSLAETLTOAQERGEDMDGCVLCRVQQ
RROSDREATIGVDSLVHLFANRWAPLVVGRNIGLMTWELTFPARDVLAORTLGNVAR"
/complement(7749..9599)
/notes="similar to GenBank Accession Number D00398
(ECOAPP2)"
gene
complement(7906..9231)
CDS
/gene="pepp"
complement(7906..9231)
/notes="ORF_f441; third start codon"
/codon_start=1
/transl_table=11
/protein_id="AAAG69076.1"
/db_xref="GI:882438"
/translation="MSEISRQFRRRQALVEQMPGSAALIFAAPEVTRSADSEYPY
RQNSDFWYFTGFNEPEAVLVLIKSDDTNHNHSLVFNVRDLTAEIFGRRLGQDAAPEK
LGVDRALAFSEIQVOLLLGDVYVHAQGEYADYIVNSALEKLRKGSRONLTAP
ATWIDRPPVVMRLPKSPREIAVLRAGEITAMATRAKEKCPGMEFYHLEGEIHH
FNRHGRAPSYNTIIVSGSGCILLHYTENECMRDGLVLLIDAGCEIKGYAGDITRT
FPVNGFTQAKREIYDILVLESLETSRLRYPGTSILEVTGEVVRIMVSGLVKLGILKG
DVDELIAQNAHRPFHGLSHLGDVHDVYVGYQDRSRIEPLPEGMVLTPEPGLYIAPD
AEVPEQYIGIRIEDDITVITGTGNNENLTASVYVKPEETIEALMVAARKQ"
/gene="ygfB"
complement(9257..9841)
CDS
/gene="ygfB"
/notes="ORF_f194"
/codon_start=1
/transl_table=11
/protein_id="AAAG69077.1"
/db_xref="GI:882439"
/translation="MLMSIQNMPGYNEMNQYLNOQGTGLTPAEMHGLISGMTCGGND
DSSWLPLLDLTNEGMAFGHELAAQAKVHSATSDALODDGFULYLPDGDVSVFD
RADALAGWVNHLLGLGVTPQKLDKVTGETGEAIDDLRNIAQLGYDEDEQOELENSL
EEIEYVRVAALLCHDITFTHPQTAPEVOKPTLH"
/10003..10332
/notes="ORF_o109"
```

```
misc_feature
10266..11219
/notes="similar to GenBank Accession Number M12965
(ECOSSR)"
misc_RNA
10374..10556
/gene="ssr"
/product="stable 6S RNA"
gene
10374..10556
/gene="ssr"
misc_feature
10374..10557
/notes="similar to GenBank Accession Number X01238
(ECRRN6S)"

Query Match      8.58; Score 60.4; DB 2; Length 141744;
Best Local Similarity 51.18; Pred. No. 7.1e-07;
Matches 142; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 49 tccacctgtttgtcattggtcttcttcgctgcatgacgcaacaccataacagtaagcg 108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22910 TCAATATTACTGCTCATTTGCTGTTTTAAAGCGCCAAACCATACACAGCAAGAGGCT 22969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 109 atgccacccagcgcaataatcagcagcagtcgcgcgcgccattggacaaaagatcaacy 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22970 ACGCCGCCCGCAGGTGACCGCGCCAGCAACCCAGCGGCTGCATCAATAACGGCGTCCCA 23029
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 169 cccaaagtgcggcgatgaacaaagacgctcagaaattaaacacacagagaagaccgca 228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23030 CCAAAAATCCCGCGCCAAATCAGGACCAAAATCGCTGATAGCACAAGTAAGGCAATCATA 23089
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 229 atgagtccttcgagcttaattccttctgttttaaccacagatcattcgcgtccgatggac 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23090 ATGTGGTACTGACGACGATGCTGCTGCTGATTCATCACAAGCAATTTTGTGGACCGCGGT 23149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 289 agtaaaagactggcccccaaaagcagacctgtaataa 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23150 AGGATCATAGCCGCCCAAGTGCAAGACCTTGAAGTA 23187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: September 14, 2001, 07:52:13
Job time: 15035 sec
```



Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	63.2	8.9	15239	1	US-08-390-878-17		Sequence 17, Appl
2	37	5.2	2058	1	US-08-358-117-1		Sequence 1, Appl
3	37	5.2	2058	3	US-08-470-588-1		Sequence 1, Appl
4	33	4.6	2656	2	US-08-685-625A-5		Sequence 5, Appl
5	32.8	4.6	1876	3	US-08-714-918-33		Sequence 33, Appl
6	32.8	4.6	1876	4	US-09-265-315-33		Sequence 33, Appl
7	32.8	4.6	1876	4	US-09-268-315-33		Sequence 33, Appl
8	32.8	4.6	1876	4	US-09-266-417-33		Sequence 33, Appl
9	31.2	4.4	219	3	US-08-078-271B-9		Sequence 9, Appl
10	31	4.4	4700	4	US-09-150-460B-9		Sequence 9, Appl
11	31	4.4	7218	1	US-08-232-463-14		Sequence 14, Appl
12	30.8	4.3	8501	3	US-08-793-900-1		Sequence 1, Appl
13	30.6	4.3	1299	4	US-09-385-028-21		Sequence 21, Appl
14	30.6	4.3	11604	4	US-09-385-028-13		Sequence 13, Appl
15	30.6	4.3	15079	4	US-09-385-028-1		Sequence 1, Appl
16	30.2	4.2	9207	3	US-08-388-353-800		Sequence 800, App
17	30.2	4.2	9207	3	US-08-488-515B-800		Sequence 800, App
18	30	4.2	213	1	US-07-867-194-3		Sequence 3, Appl
19	30	4.2	259	1	US-07-867-194-4		Sequence 3, Appl
20	30	4.2	261	1	US-07-867-194-5		Sequence 5, Appl
21	29.6	4.2	1872	1	US-08-428-926-1		Sequence 1, Appl
22	29.6	4.2	1872	1	US-08-435-434-4		Sequence 4, Appl
23	29.6	4.2	1872	1	US-08-435-436-4		Sequence 4, Appl
24	29.6	4.2	1872	1	US-08-428-927-1		Sequence 1, Appl
25	29.6	4.2	1872	1	US-08-428-298-1		Sequence 1, Appl
26	29.6	4.2	1872	2	US-08-339-517-1		Sequence 1, Appl
27	29.6	4.2	1872	2	US-08-438-863-4		Sequence 4, Appl



```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
SUITE: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,918
FILING DATE: September 13, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 1876 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-714-918-33

Query Match          4.6%; Score 32.8; DB 3; Length 1876;
Best Local Similarity 48.3%; Pred.No. 0.77; Indels      0; Gaps
Matches    70; Conservative   7; Mismatches    68;

QY  292 aaagactggccccaagaacgacacctgaatgaagatttccatcaccatcgacc 351
     ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB  606 ARAHTAATTTCTTCADCCAHAGAAGGATGGATGATTGTGCWTCATAATGAAG 665
     |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |

QY  352 tatggaagtacttaadgaataatgatgttgctttaacatggtttaatatagcttcatac 411
     |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
DB  666 AATGGTITTCAGATATKCGACAGTTTTGGTTWTWTACCCTGGATTATATGTTATATTA 725
     |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |

QY  412 ccattcaactggacaacttgtct 436
     ||| | | ||| ||| |||
DB  726 GAGCTNTAAATCCAGTATTTTCCTTT 750
     |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |

RESULT        6
US-09-265-315-33
Sequence 33, Application US/09265315
Patent No. 6187541
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ving J.
APPLICANT: Malouin, Francis
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
```

ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
SUITE: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/265,315  
FILING DATE: March 9, 1999  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/714,918  
FILING DATE: September 13, 1996  
APPLICATION NUMBER: 60/009,102  
FILING DATE: December 22, 1995  
APPLICATION NUMBER: 60/003,798  
FILING DATE: September 15, 1995

ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 240/247

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510

INFORMATION FOR SEQ. ID NO.: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1876 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-09-265-315-33

[illegible]

```

RESULT      7
US-09-265-315-33
: Sequence 33, Application US/09265315
: Patent No. 6187541
: GENERAL INFORMATION:
: APPLICANT: Benton, Bret
: APPLICANT: Lee, Ving J.
: APPLICANT: Malouin, Francois
: APPLICANT: Martin, Patrick K.
: APPLICANT: Schmid, Molly B.
: APPLICANT: Sun, Dongxu
: TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
: ACTIVE ON STAPHYLOCOCCUS AUREUS
: TITLE OF INVENTION: TARGET GENES
: NUMBER OF SEQUENCES: 111

```

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/265,315  
FILING DATE: March 9, 1999  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/714,918  
FILING DATE: September 13, 1996  
APPLICATION NUMBER: 60/009,102  
FILING DATE: December 22, 1995  
APPLICATION NUMBER: 60/003,798  
FILING DATE: September 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 240/247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1876 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-265-315-33

	Query Match	4.6%	Score 32.8;	DB 4;	Length 1876;
	Best Local Similarity	48.3%;	Pred. No. 0.77;	Mismatches	0;
	Matches	70;	Conservative	7;	Gaps
QY	292	aaaagactggccccaaaaagcagaccctgttaatgaagatttccatgatcacacatcgtagacc	351		
		: :      : :  : :  : :  : :  : :  : :  :			
Dd	606	ARAHTAATTAFTCTTCACCAHAGAAAGGATTGATGATTTCGTCTWTMTCAATTAGAAG	665		
		: :  : :  : :  : :  : :  : :  : :  : :  : :  :			
QY	352	tatggaagtacctaaagtataatgatggttcctaacaatggtttaatatagcttcacatgaac	411		
		: :  : :  : :  : :  : :  : :  : :  : :  : :  :			
Dd	666	ANTCGTTTGAGATAATKCGACAGTITGGTWTTTACCTGGATTATATGTTATATATTA	725		
		: :  : :  : :  : :  : :  : :  : :  : :  : :  :			
QY	412	cccattcaactgggcacatttgtctct	436		
		: :  : :  : :  : :  : :  : :  : :  : :  : :  :			
Dd	726	GAGCNTNTAAATCCAGTATTTCCTTT	750		
		: :  : :  : :  : :  : :  : :  : :  : :  : :  :			

```

RESULT      8
US-09-266-417-33
; Sequence 33, Application US/09266417
; Patent No. 6228588
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving J.
; APPLICANT: Marouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
; TITLE OF INVENTION: TARGET GENES

```

```

: NUMBER OF SEQUENCES: 111
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071-2066
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: MEDIUM TYPE: storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: Word Perfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/266,417
: FILING DATE: March 9, 1999
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/714,918
: FILING DATE: September 13, 1996
: APPLICATION NUMBER: 60/009,102
: FILING DATE: December 22, 1995
: APPLICATION NUMBER: 60/003,798
: FILING DATE: September 15, 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 240/248
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 33:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1876 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-266-417-33

```

Query Match 4.6%; Score 32.8; DB 4; Length 1876;  
Best Local Similarity 48.3%; Pred. NO. 0.77;  
Matches 70; Conservative 7; Mismatches 68; Indels 0

Qy	292	aaagactggcccccaaaagcagacatgtaataagaattcccatgatcaccatctgtacc	351
		:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	606	ARAHATAATTCTTCADCCAHAGAAAGGATTCATGATTGTGCGTWCMTCAATTAGAAG	665
Qy	352	tatggaagtacttaaggaaaatgattgggttcttaacaatggttttaatatagcttcataaac	411
		:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	666	AATGGTTTGAGATATKCGACACAGTTGGTTWTTTACCCTGGATTTATATTGTTATATATTA	725
Qy	412	ccattcaactggacaccttttgtctct	436
		:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	726	GAGCTNTAATTCAGTAGTATTTCCCTTT	750

RESULT 9  
US-08-078-271B-9  
; Sequence 9, Application US/08078271B  
; Patent No. 6136527  
; GENERAL INFORMATION:  
; APPLICANT: Fuchs, Klaus; Motz, Manfred; Roggendorf, Michael;  
; APPLICANT: Soutschek, Erwin  
; TITLE OF INVENTION: Polypeptides Derived from Proteins of the  
; TITLE OF INVENTION: Hepatitis C Virus, Test Kits Containing these Polypeptides  
; TITLE OF INVENTION: Hepatitis C Virus, Vaccines against Infection by Hepatitis C Virus  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felpe & Lvnch

STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/078, 271B  
FILING DATE: 15-OCT-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DE91/01020  
FILING DATE: 20-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 40 41 304.7  
FILING DATE: 21-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 6136527man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LEADER 201  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

95-08-078-271B-9

Query Match 4.4%; Score 31.2; DB 3; Length 219;  
Best Local Similarity 50.7%; Pred. No. 0.85;  
Matches 75; Conservative 0; Mismatches 73; Indels

y	105	agcgtatgccacccaccgcgcatataatatcgagcacgatcgcgcgcatctggdacaagaatc	164
b	31	AGGGGGCCCTGCGGAGGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATC	90
y	165	aacgcccaagtgcgcgcgatgaacaaaagacgcagaaattaaacacacgagaagaac	224
b	91	ACCGGACATGTCAGAACGGGACGATGAGGATCTCGTGCTTAAGACCTCGAGGAACATG	150
y	225	cgaatatgagtaccttcgcgccttaattcct	252
b	151	TGGAGTGGGACCTTCCCATTAAATGCCT	178

```

RESULT 10
US-09-150-460B-9
; Sequence 9, Application US/09150460B
; Patent No. 6190882
; GENERAL INFORMATION:
; APPLICANT: Lee, Cheng-Chi
; APPLICANT: Aldrecht, Urs
; APPLICANT: Eichele, Gregor
; APPLICANT: Sun, Zhong Sheng
; TITLE OF INVENTION: Mammalian Circadian Rhythm-Like Gene
; FILE REFERENCE: D6039
; CURRENT APPLICATION NUMBER: US/09/150,460B
; CURRENT FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: US 60/058,256
; PRIOR FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 9
; LENGTH: 4700
; TYPE: DNA
; ORGANISM: artificial sequence

```

```

; FEATURE:
; OTHER INFORMATION: Nucleotide sequence of m-rigul, a murine brain cDNA
;
; OTHER INFORMATION: homologous to a human RIGUI 4.7 cDNA probe
US-09-150-460B-9

```

	Query Match	4.4%	Score 31;	DB 4;	Length 4700;
	Best Local Similarity	49.7%;	Pred. No. 4.7;		
	Matches 79;	Conservative 0;	Mismatches 80;	Indels 0;	Gaps 0;
QY	451	ggcagcttcgaagcgcgctctcttagccctttccatttccccctcgcggtgaagtcagcgc	510		
Db	3057	gccaccacgctgtcttcccactgcctctccactctgccccaccacctctcagccc	3116		
QY	511	gttaaacgtcttcgagcatcacgtgggtcgagtgttgatcgcgaccaccaacggcgcaaa	570		
Db	3117	ccccaccgcccagatccccacgttcaactcagatcgagctcccccactccagctcaa	3176		
QY	571	gcaaccgaagcgggtgaagtcttgtgcaagcagcgcg	609		
Db	3177	tctcgtcagctttagagtgctcccccgcacgagaggggg	3215		

```

RESULT 11
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
;

```

```

Query Match      4.4%; Score 31; DB 1; Length 7218;
Best Local Similarity 14.2%; Pred.No. 5.8;
Matches 52; Conservative 140; Mismatches 175; Indels 0; Gaps

QY   84  gacgccaaccataacaggtgaagcgcgatgcccccagcgcataatctcgagcacgacgcg 143
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   1310 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY   144  cgcggcattggacaaaagatcaacgccccaggctgcgcggcgatgaacaaaaagaagtcaga 203
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   1250 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY   204  aattaacaacacagagaagaacgcgaalgaagtctctcgogcttaattcttctttaatacc 263
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   1190 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY   264  cagtacattctgcggtccgatgacagtgaaagactgcccccaaacagacacctgttaat 323
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   1130 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY   324  gaagattttccatgatcaccatgctgacctatggaagtacttaagtaaatgattggttct 383
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   1070 RRRRTATCGCAAGCTCCCTCGACCTGCAGCCAGCTCGGAATTAACTCTGAGCGTATGG 1011
QY   384  taacatggtttaatatagttctcatgaacccccattcaactggacaccttgcctctaatac 443
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   1010 CAAAAGAAAGAAAAATAGTTATAGTAGCGCGACTCGATGGGACATTTCAAAGCTAAACGT 951

QY   444  tgaatga 450
DB   950  TTAATAA 944

RESULT 12
US-08-793-900-1/c
; Sequence 1, Application US/08793900
; Patent No. 6143518
; GENERAL INFORMATION:
; APPLICANT: CAMERON, Beatrice
; APPLICANT: CROUZET, Joel
; TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF RECOMBINANT
; TITLE OF INVENTION: PROTEINS, PLASMIDS AND MODIFIED CELLS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Road, Mailstop 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,900
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR95/01178
; FILING DATE: 14-SEP-1995
; APPLICATION NUMBER: FR 94/11049
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Savitzky Esq., Martin F.
; REGISTRATION NUMBER: 29,699
; REFERENCE/DOCKET NUMBER: ST94069-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 454-3816
; TELEFAX: (610) 454-3808
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```



LENGTH: 8501 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: cDNA  
US-08-793-900-1

Query Match 4.3%; Score 30.8; DB 3; Length 8501;  
Best Local Similarity 53.3%; Pred. No. 7.4;  
Matches 65; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
QY 508 cgcgttaagctctcagacatcacgtggctcagagtggtggtatcgacaccccaacggcc 567  
DB 6766 CGGGCCAAAGGGCCCAAGCTGGCGCTGCTGGCTGGACCTGTGGAGCTGGCGGCC 6707  
QY 568 aaagcaacacgagcgggtgaagctctgtcgaagcagcgcgaaatggtgtgctgcaa 627  
DB 6706 GAGCGGACGAGGATGGGGAATAATCGTCTGGAATCCGTCAGGACATGCTTTGAAGCTC 6647  
QY 628 gc 629  
DB 6646 GC 6645

RESULT 13  
US-09-385-028-21/c  
Sequence 21, Application US/09385028  
Patent No. 6232106  
GENERAL INFORMATION:  
APPLICANT: Susan E. Jensen  
APPLICANT: Kwamena A Aidoo  
APPLICANT: Ashish S. Paradkar  
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
Patent No. 6232106  
TITLE OF INVENTION: Acid Biosynthesis  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PPLC  
STREET: The Jennifer Building, 400 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/385,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/790,462  
FILING DATE: 29-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: D. Douglas Price  
REGISTRATION NUMBER: 24,514  
REFERENCE/DOCKET NUMBER: 1418/P57452US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638-6666  
TELEFAX: (202) 39305350  
TELEX: RCA 248593 IDEA UR  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1299 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-385-028-21

Query Match 4.3%; Score 30.6; DB 4; Length 1299;  
Best Local Similarity 53.8%; Pred. No. 3.3;  
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
QY 449 aaggcagcttcgaagcgccctccttagccctttccattccctccctcgccggtgagtcagc 508  
DB 284 ACGGACGATCAACCGGGCGCGGGTGTGTCGATCAGCCAGTCGGCGCTCACC CGCC 225  
QY 509 gcgttaagctctcagacatcacgtggctcagtgagtggtgtggtatcgacaccccaacgg 565  
DB 224 GGATTCACCGCTGGAGCAGTCCTCGCGCGCGGCTGCTGCTGCGAGCCCGCGG 168

RESULT 14  
US-09-385-028-13/c  
Sequence 13, Application US/09385028  
Patent No. 6232106  
GENERAL INFORMATION:  
APPLICANT: Susan E. Jensen  
APPLICANT: Kwamena A Aidoo  
APPLICANT: Ashish S. Paradkar  
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
Patent No. 6232106  
TITLE OF INVENTION: Acid Biosynthesis  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PPLC  
STREET: The Jennifer Building, 400 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/385,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/790,462  
FILING DATE: 29-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: D. Douglas Price  
REGISTRATION NUMBER: 24,514  
REFERENCE/DOCKET NUMBER: 1418/P57452US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638-6666  
TELEFAX: (202) 39305350  
TELEX: RCA 248593 IDEA UR  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11604 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-385-028-13

Query Match 4.3%; Score 30.6; DB 4; Length 11604;  
Best Local Similarity 53.8%; Pred. No. 10;  
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
QY 449 aaggcagcttcgaagcgccctccttagccctttccattccctccctcgccggtgagtcagc 508  
DB 9249 ACGGACGATCAACCGGGCGCGGGTGTGTCGATCAGCCAGTCGGCGCTCACC CGCC 9190  
QY 509 gcgttaagctctcagacatcacgtggctcagtgagtggtgtggtatcgacaccccaacgg 565

Search completed: September 14, 2001, 07:51:13  
Job time: 12387 sec



117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

RESULT	1	
CNS01MQH		
LOCUS	830 bp	DNA
DEFINITION	Anopheles gambiae BAC end sequence T7 end of clone 22E24 of NotreDamel library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.	
ACCESSION	AL151258	
VERSION	AL151258.1	GI:7011737
KEYWORDS	GSS.	
SOURCE	African malaria mosquito.	
ORGANISM	Anopheles gambiae	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
	Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;	
	Culicoidae; Anopheles.	
REFERENCE	1 (bases 1 to 830)	
AUTHORS	Genoscope.	
TITLE	Direct Submission	



BASE COUNT







```

QY 481 tccattcccccctcgcg 499
    ||| | ||| |||
Db 248 TCCACGACACCTTGGCAG 266

RESULT 11
AW351850/c 166 bp mRNA EST 01-FEB-2000
DEFINITION RC1-CT0199-180999-012-A09 CT0199 Homo sapiens cdna, mRNA sequence.
ACCESSION AW351850
VERSION AW351850.1 GI:6849563
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC1&t2=RC1-CT0199-
180999-012-A09&t3=1999-09-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 51
High quality sequence stop: 166.
Location/Qualifiers
  1..166
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="CT0199"
    /dev_stage="Adult"
    /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
    SmaI; A mini-library was made by cloning products derived
    from ORESTES PCR (U.S. Letters Patent application No. 196
    ,716 - Ludwig Institute for Cancer Research) profiles
    into the pUC 18 vector. Reverse transcription of tissue
    mRNA and cDNA amplification were performed under low
    stringency conditions."
BASE COUNT 42 a 41 c 52 g 31 t
ORIGIN

Query Match 5.0%; Score 35.6; DB 115; Length 166;
Best Local Similarity 58.5%; Pred. No. 3.3;
Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 12 caccgtgtgtcttctcaatgatctgtgcgcttcacacctgttctgtcatgctc 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137 CACATGGGTGGTGTCTCTACAAATGGTCACAGCCTTACACCTCGTTCTTTCGCTT 78

QY 72 ttctcgtgcatacagcgaaaccataacacaggttaagcgatgccacc 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 GGATCCCGCCCTGTGACATTCGCCGACCATGTGAGTTATGCCAGCC 32

RESULT 12
BF352878 222 bp mRNA EST 22-NOV-2000
DEFINITION IL3-HT0619-100700-208-H08 HT0619 Homo sapiens cdna, mRNA sequence.
ACCESSION BF352878
VERSION BF352878.1 GI:11311952
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 222)
AUTHORS Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL3&t2=IL3-HT0619-
100700-208-H08&t3=2000-07-10&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 60
High quality sequence stop: 91.
Location/Qualifiers
  1..222
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="HT0619"
    /dev_stage="Adult"
    /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
    Site_2: SmaI; A mini-library was made by cloning products
    derived from ORESTES PCR (U.S. Letters Patent application
    No. 196,716 - Ludwig Institute for Cancer Research)
    profiles into the pUC 18 vector. Reverse transcription of
    tissue mRNA and cDNA amplification were performed under
    low stringency conditions."
BASE COUNT 45 a 67 c 55 g 55 t
ORIGIN

Query Match 5.0%; Score 35.6; DB 147; Length 222;
Best Local Similarity 58.5%; Pred. No. 3.5;
Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 12 caccgtgtgtcttctcaatgatctgtgcgcttcacacctgttctgtcatgctc 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 CACCATGGGTGGTGTCTCTACAAATGGTCACAGCCTTCCACACCTTCTTCTTTCACCTT 64

QY 72 ttctcgtgcatacagcgaaaccataacacaggttaagcgatgccacc 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 GGATCCCGCCCTGTGACATTCGCCGACCATGAGTGAGTCATGCCACCC 110

RESULT 13
AQ851405 393 bp DNA GSS 13-APR-2001
LOCUS LMAJFV1_lm40b11.y1 Leishmania major FV1 random genomic library
DEFINITION Leishmania major genomic clone LMAJFV1_lm40b11 5', DNA sequence.
ACCESSION AQ851405
VERSION AQ851405.1 GI:6056053
KEYWORDS GSS.
SOURCE Leishmania major.
ORGANISM Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 393)
AUTHORS Akopyants,N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L.,

```

Kissinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A., Blistain, A., Schmitt, A., Person, B., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M., Harvey, N., McCann, R., Tsagaris, R., Williams, T., Jackson, Y., Bowers, Y., Swaller, T., Waterston, R., Wilson, R., and Beverley, S.M.  
A survey of the Leishmania major Friedlin strain V1 genome by shotgun sequencing: a resource for DNA  
Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)  
Other\_GSSs: lm40b11.x1  
Contact: Akopyants, NS / Beverley, SM  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library construction: Natalia S. Akopyants, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
If using this information please cite:  
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major Friedlin strain V1 genome by shotgun sequencing' and the Washington University Genome Sequencing Center For information on obtaining clone material please contact: Natalia S. Akopyants Ph.D. (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D. (beverley@borcim.wustl.edu)  
Seq primer: -40RP from Gibco  
Class: shotgun  
High quality sequence stop: 357.  
Location/Qualifiers  
source  
1. .393  
/organism="Leishmania major"  
/strain="Friedlin strain V1"  
/db\_xref="taxon:5664"  
/clones="LMAJFV1\_lm40b11"  
/lab\_host="TOP10 (Invitrogen)"  
/note="vector: pZero-2 (Invitrogen); Site\_1: EcoRV;  
Genomic DNA was isolated from stationary phase cells. For this library, DNA was sheared to give a tight size distribution of 1-1.5kb fragments, blunt-ended with T4 DNA polymerase, dephosphorylated with Shrimp Alkaline Phosphatase and ligated into pZero-2 vector's EcoRV site."  
BASE COUNT 105 a 97 c 116 g 75 t  
ORIGIN  
Query Match 5.0%; Score 35.6; DB 234; Length 393;  
Best Local Similarity 53.6%; Pred. No. 4.2;  
Matches 74; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
Qy 564 ggcacaaagcaccgagcggtgaagtccttctgtgcaagcagcggaataatggtgtgct 623  
Db 185 GTCCGTAGCGCAGTGTGCGGAAACCTCACCCCTCAAGCAGCGAGTTGCTGGTGGGCT 244  
Qy 624 gcaagcagaactaaagcgcaactatcttgagcgcttctgtaaatcccggttaaccatgc 683  
Db 245 ACTAGCAGTAATCGCAGCAGCGAAAGCGGAGGATCTGTAAGACCTGTTAGCCACGGC 304  
Qy 684 catcaacgacattcgct 701  
Db 305 ACTAAGCGGAATCTGAT 322  
RESULT 14  
AW067177/c  
LOCUS AW067177 492 bp mRNA EST 12-OCT-1999  
DEFINITION 683016G08.x1 683 - 14 day immature embryo from Hake lab (HS) Zea mays cDNA, mRNA sequence.  
ACCESSION AW067177  
VERSION AW067177.1 GI:6022249  
KEYWORDS EST.  
SOURCE Zea mays.  
ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.  
1 (bases 1 to 492)  
AUTHORS Walbot, V.  
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University  
JOURNAL Unpublished (1999)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 683016 row: G column: 08.  
Location/Qualifiers  
source  
1. .492  
/organism="Zea mays"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone\_lib="683 - 14 day immature embryo from Hake lab (HS)"  
/tissue\_type="embryo"  
/dev\_stage="14 days after pollination"  
/lab\_host="DH10B"  
/note="Organ: embryo; Vector: pBKCMV (Stratagene's Zap Express); Site\_1: XhoI; Site\_2: EcoRI; Directionally cloned, 14 day immature embryo library created with Stratagene's Zap Express cDNA protocol."  
BASE COUNT 125 a 97 c 154 g 116 t  
ORIGIN  
Query Match 5.0%; Score 35.6; DB 111; Length 492;  
Best Local Similarity 51.9%; Pred. No. 4.4;  
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;  
Qy 231 gagtcttcgcgcttaattcctgttttaaccacgactatctgcgctcgagacag 290  
Db 399 GAGTCCCTCGTGTGTGCCCTTTTGTGCCCCGAGGGTGGCTGCTTTTCAAGTTCCTG 340  
Qy 291 taaagactggcccccaagcagacctgttaagaaagatttcctcatcaccatcgtagc 350  
Db 339 TAAAGGGGACAAACCCCAAGCCCGCAGAGTATACGTTTCCCGCCGCCGTGGCCA 280  
Qy 351 ctatggaagtacttaagtaaaatgattgttctt 384  
Db 279 ATATCGATATACTCTGCGCAATCGCCAGTTCCT 246  
RESULT 15  
AZ253065  
LOCUS AZ253065 494 bp DNA GSS 15-JUN-2000  
DEFINITION RPCI-23-94A11.TV RPCI-23 Mus musculus genomic clone RPCI-23-94A11, DNA sequence.  
ACCESSION AZ253065  
VERSION AZ253065.1 GI:8566268  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 494)  
AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C.M.  
TITLE Mouse BAC End Sequences from Library RPCI-23  
JOURNAL Unpublished (1999)  
COMMENT Other\_GSSs: RPCI-23-94A11.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhae@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 94 row: A column: 11

Seq primer: T7  
 Class: BAC ends.  
 Location/Qualifiers  
 1. .494  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="RPCI-23-94A11"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 220 a 71 c 73 g 130 t  
 ORIGIN

Query Match 5.0%; Score 35.6; DB 240; Length 494;  
 Best Local Similarity 50.6%; Pred. NO. 4.4;  
 Matches 86; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
 QY 241 cgcttaattccttggttaataccagatcacattctcggtgcgagtgacagtaaaagactg 300  
 Db 318 CTCTAAACATTTTAAAAAAGTCTATTTCCTCCCAATGAAGATGATAAATCCACCT 377  
 QY 301 gcccccaaaagacacgtgaatgaagatttccatgataccatcgtgacctatggaagt 360  
 Db 378 AATCCTATAACAAAGACCAACAATGATTACCATGCTTGACATAATGATAGTCAAAAT 437  
 QY 361 actaagtaaatgattggttcttaacatggtttatagcttcattgaa 410  
 Db 438 TATATACAAAATANTTATGCAATACCTAAATCAAAATCACTGGAGAA 487

Search completed: September 14, 2001, 06:59:12  
 Job time: 12059 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2001, 05:26:31 ; Search time 171.99 Seconds  
(without alignments)  
2595.721 Million cell updates/sec

Title: US-09-105-117i-1-copy\_1016\_1726  
Perfect score: 711  
Sequence: 1 ggtcatcgggcacggttgg.....cagattcgctatccacatgg 711

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0601:\*

1:	/cgnl_9/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/cgnl_9/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/cgnl_9/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/cgnl_9/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/cgnl_9/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/cgnl_9/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/cgnl_9/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/cgnl_9/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/cgnl_9/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/cgnl_9/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/cgnl_9/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/cgnl_9/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/cgnl_9/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/cgnl_9/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/cgnl_9/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/cgnl_9/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/cgnl_9/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/cgnl_9/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/cgnl_9/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/cgnl_9/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/cgnl_9/gcgdata/geneseq/geneseq/NA2000.DAT:*
22:	/cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	711	100.0	2374	18	AA96816
C 2	435	61.2	822	22	AA971779
C 3	406	57.1	993	22	AA971777
C 4	63.2	8.9	15239	17	AA933536
C 5	60.4	8.5	636	21	AA52891
C 6	45.4	6.4	45613	22	AA928535
C 7	45.2	6.4	4915	20	AA984317
C 8	37.8	5.3	936	22	AA958252
C 9	37.8	5.3	936	22	AA958254
C 10	37.8	5.3	936	22	AA958257
C 11	37.8	5.3	936	22	AA958259

12	37.8	5.3	936	22	AA958262	Oligonucleotide D2
13	37.8	5.3	936	22	AA958255	Oligonucleotide D1
C 14	37	5.2	2058	9	AA980093	Sequence of BamHI/
C 15	37	5.2	2058	16	AA97857	P. aeruginosa ttda
C 16	37	5.2	102634	21	AA981464	N. meningitidis pa
C 17	37	5.2	349980	21	AA921609	Neisseria meningit
C 18	36.6	5.1	10732	21	AA910594	Gene encoding a su
C 19	36	5.1	936	22	AA958252	Oligonucleotide D1
C 20	36	5.1	936	22	AA958254	Oligonucleotide D1
C 21	36	5.1	936	22	AA958257	Oligonucleotide D1
C 22	36	5.1	936	22	AA958259	Oligonucleotide D2
C 23	36	5.1	936	22	AA958262	Oligonucleotide D2
C 24	36	5.1	936	22	AA958255	Oligonucleotide D1
C 25	34.8	4.9	1431	8	AA970284	Sequence encoding
C 26	34.6	4.9	245	21	AA922623	Human secreted pro
C 27	34.6	4.9	1704	20	AA996998	Human TGF-beta act
C 28	34.6	4.9	2785	20	AA996996	Human TGF-beta act
C 29	34.6	4.9	2866	20	AA996997	Human TGF-beta act
C 30	34.4	4.8	5819	20	AA920533	Polynucleotide seq
C 31	33	4.6	1788	20	AA956285	Human TAK1-6xHis e
C 32	33	4.6	1959	18	AA985095	Human transforming
C 33	33	4.6	2656	20	AA956279	Human TAK1 encodin
C 34	33	4.6	2656	21	AA939105	Human TAK-1 nucleo
C 35	32.8	4.6	1876	21	AA926882	Essential Staphylo
C 36	32.8	4.6	1876	22	AA91564	Staphylococcus aur
C 37	32.6	4.6	346	21	AA943240	Xenopus secreted e
C 38	32.6	4.6	1134	21	AA943255	Neisseria meningit
C 39	32.6	4.6	1134	21	AA94489	Neisseria meningit
C 40	32.6	4.6	25509	21	AA981467	N. meningitidis pa
C 41	32.6	4.6	349980	21	AA921607	Neisseria meningit
C 42	32.4	4.6	3333	21	AA93618	Human semaphorin p
C 43	32.4	4.6	3498	21	AA93617	Human semaphorin p
C 44	32.4	4.6	3550	19	AA944295	Human secreted pro
C 45	32.4	4.6	3862	21	AA901233	Human semaphorin 6

ALIGNMENTS

RESULT 1  
AA96816/c  
ID AA96816 standard; DNA; 2374 BP.  
XX AA96816;  
DT 12-MAR-1998 (first entry)  
XX  
DE DNA encoding LysG, LysE and ORF3 from Corynebacterium glutamicum.  
XX  
LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;  
KW Microbial production; amino acid; animal feed additive; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
FH Key Location/Qualifiers  
CDS complement (82..954)  
FT /\*tag= a  
FT /label= LysG  
FT 1016..1726  
FT /\*tag= b  
FT /label= LysE  
FT complement (1723..2373)  
FT /\*tag= c  
FT /label= orf3  
DE19548222-A1.  
XX  
PD 26-JUN-1997.  
XX  
PF 22-DEC-1995; 95DE-1048222.  
XX  
PR 22-DEC-1995; 95DE-1048222.  
XX

PA (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.

XX Eggeling L, Sahm H, Vrljic M;

XX WPI; 1997-333867/31.

DR P-PSDB; AAW37714-16.

XX

PT Increasing microbial production of amino acids, especially lysine -  
PT by improving export carrier activity or corresponding gene  
PT expression, also new export and regulatory genes from  
PT Corynebacterium

XX Claim 23 and 26; Page -: 16pp; German.

XX This DNA, isolated from Corynebacterium glutamicum, contains the LysG,  
CC LysE and Olf3 genes. LysG and LysE encode a lysine transport regulatory  
CC protein and an export protein, respectively. Microbial production of  
CC amino acids (A) is improved by increasing the export-carrier activity  
CC and/or the export gene expression in a microorganism that produces (A).  
CC The method is specifically used to increase production of lysine,  
CC used as an animal feed additive. Other (A) are variously useful as  
CC pharmaceuticals, condiments and intermediates for fine chemicals.  
CC This method increases the amount of (A) secreted into the culture medium.  
CC Export of (A) has been found to depend on a single gene.  
CC NB. This sequence has been created from the information given in table 2  
CC of the specification.

XX Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 711; DB 18; Length 2374;

Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcatcgaggcaggttggtctgttttttcaatgatgtgtggtccacccctgttt 60

Db 1359 GTGTCAATCGAGCAGCTGCGCGCGGATTCGACAAAGATCAACGCCAAGGTGCGG 1300

QY 61 gtcatggtcttcttcgtccatgacggcaaacataacaggttaagcgtgccaccccg 120

Db 1299 GTCATGGCGTCTTCGCTGCCATGACGGCAACCATACAGGTAAAGCGATGCCACCCAG 1240

QY 121 cgataatctgagcagctcgcgcggttgacaaagatgacacgcccgaagtgccg 180

Db 1239 CGCATTAATTCGAGCAGCTGCGCGCGGATTCGACAAAGATCAACGCCAAGGTGCGG 1180

QY 181 gcatgaacaaagacgctcaaaaataaacacacagagaaaccccaatgagctctcg 240

Db 1179 GCATGAACAAAGACGTCAGAAATTAACACACGAGAGAACCCGCAATGAGTCTCTCG 1120

QY 241 cgcttaattctctgttcaatcaccagttacattctgcggtccggtgacacagtaaaagactg 300

Db 1119 CGCTTAATTCCTGTGTTAATCACCAGTACATCTCGCGTCCGATGGACAGTAAAGACTG 1060

QY 301 gcccccaagacagacctgaatgaagatttcacatgataccatcgtgacctatggaagt 360

Db 1059 GCCCCCAAGACAGACCTGAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGT 1000

QY 361 acttaagtaaaatgattgtttcttaacatggttttaatatagcttaagccccattcaa 420

Db 999 ACTTAAGTAAATGATGTGTTCTTAACATGGTTTAAATAGCTTCATGAACCCCAATCAA 940

QY 421 ctggacactttgtctcaatcatctgatgagcgagcttcgaagcgccctccttagccctt 480

Db 939 CTGGACACTTGTCTCTCATCATGATGATGAAGCAGCTTCGAAGCGCCTCTTAGCCCTT 880

QY 481 tcatctccccctcggtgagtcgagcgtttaagctctcagacatcacgtgggtcga 540

Db 879 TCATTTCCCTCCGCGGTGAGTCAGCGGTAAAGCTCTCAGCATCACGTGGGTGCA 820

QY 541 gtgttggtatcgacaccccaacggcgaacgaacgaacgaacgaacgaacgaacgaacgaac 600

Db 819 GTGTTGTTATCGCGCACCCCAACCGGCAAGCAACCGGAGGAGTCTCTGTGCAA 760

QY 601 gcagcgcggaataatggtgtgtgcaagcagaaactaaagcgaactatcttgagcgcctt 660  
Db 759 GCAGCGCGGAAATGCTGTGTGCAAGCAGAAACTAAAGCGCAACTATCTGAGCCCTT 700  
QY 661 gctgaatacccttaaccatcgccatcgaacgagattcgctatccacatgg 711  
Db 699 GCTGAATCCCGTTAACCATCGCCATCAACGAGATTCGCTATCCACATGG 649

RESULT 2

AAF71779/c

ID AAF71779 standard; DNA; 822 BP.

XX

AC AAF71779;

XX

DT 30-APR-2001 (first entry)

XX

DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:53.

XX

Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
KW fine chemical production; microorganism; organic acid; nucleoside;  
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.

XX

OS Corynebacterium glutamicum.

XX

PN WO200100843-A2.

XX

PD 04-JAN-2001.

XX

PF 23-JUN-2000; 2000WO-IB00923.

XX

PR 25-JUN-1999; 99US-0141031.

PR

PR 01-JUL-1999; 99DE-1030476.

PR

PR 02-JUL-1999; 99US-0142101.

PR

PR 08-JUL-1999; 99DE-1031415.

PR

PR 08-JUL-1999; 99DE-1031418.

PR

PR 08-JUL-1999; 99DE-1031419.

PR

PR 08-JUL-1999; 99DE-1031420.

PR

PR 08-JUL-1999; 99DE-1031424.

PR

PR 08-JUL-1999; 99DE-1031428.

PR

PR 08-JUL-1999; 99DE-1031434.

PR

PR 08-JUL-1999; 99DE-1031435.

PR

PR 08-JUL-1999; 99DE-1031443.

PR

PR 08-JUL-1999; 99DE-1031453.

PR

PR 08-JUL-1999; 99DE-1031457.

PR

PR 08-JUL-1999; 99DE-1031465.

PR

PR 08-JUL-1999; 99DE-1031478.

PR

PR 08-JUL-1999; 99DE-1031510.

PR

PR 08-JUL-1999; 99DE-1031541.

PR

PR 08-JUL-1999; 99DE-1031573.

PR

PR 08-JUL-1999; 99DE-1031592.

PR

PR 08-JUL-1999; 99DE-1031632.

PR

PR 08-JUL-1999; 99DE-1031634.

PR

PR 08-JUL-1999; 99DE-1031636.

PR

PR 09-JUL-1999; 99DE-1032125.

PR

PR 09-JUL-1999; 99DE-1032126.

PR

PR 09-JUL-1999; 99DE-1032130.

PR

PR 09-JUL-1999; 99DE-1032186.

PR

PR 09-JUL-1999; 99DE-1032206.

PR

PR 09-JUL-1999; 99DE-1032227.

PR

PR 09-JUL-1999; 99DE-1032228.

PR

PR 09-JUL-1999; 99DE-1032229.

PR

PR 09-JUL-1999; 99DE-1032230.

PR

PR 14-JUL-1999; 99DE-1032922.

PR

PR 14-JUL-1999; 99DE-1032926.

PR

PR 14-JUL-1999; 99DE-1032928.

PR

PR 14-JUL-1999; 99DE-1033004.

PR

PR 14-JUL-1999; 99DE-1033005.

PR

PR 14-JUL-1999; 99DE-1033006.

PR

PR 12-AUG-1999; 99US-0148613.



```

PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX
PA (BADI ) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI; 2001-137957/14.
DR P-PSDB; AAB79658.
XX
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX
XX Claim 3; Page 226-228; 1737pp; English.
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.
XX
XX Sequence 993 BP; 233 A; 256 C; 268 G; 236 T; 0 other;
SQ

```

```

Query Match 57.1%; Score 406; DB 22; Length 993;
Best Local Similarity 100.0%; Pred. No. 3.1e-123;
Matches 406; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 306 caaaagcagacctgtaataagaattccatgatcaccatcgtagcctatggaagtactta 365
Db 1 caaaagcagacctgtaataagaattccatgatcaccatcgtagcctatggaagtactta 60
QY 366 agtaaaatgatgtgtcttaacatggtttaataatagatcattgaacccattcaactga 425
Db 61 agtaaaatgatgtgtcttaacatggtttaataatagatcattgaacccattcaactga 120
QY 426 cacttgtctctcaatcattgatgaagcagcttcgaagcgctctttagccttcccat 485
Db 121 cacttgtctctcaatcattgatgaagcagcttcgaagcgctctttagccttcccat 180
QY 486 ttcccctcggcggtgagtcagcggttaaaagctctcgagcatcagtggtgagtggt 545
Db 181 ttcccctcggcggtgagtcagcggttaaaagctctcgagcatcagtggtgagtggt 240
QY 546 ggtatcgccacccaaacccgcaaacgcaacggaagcggtgaagtccttgcgaagcagc 605
Db 241 ggtatcgccacccaaacccgcaaacgcaacggaagcggtgaagtccttgcgaagcagc 300
QY 606 gcgggaaatggtgtgtcgaagcagaataaaagcgcaactatctggcgcttctgta 665
Db 301 gcgggaaatggtgtgtcgaagcagaataaaagcgcaactatctggcgcttctgta 360
QY 566 aatcccggttaaccatcgccatcaacgcagattcgctatccaatgg 711

```

```

Db 361 aatcccggttaaccatcgccatcaacgcagattcgctatccaatgg 406

```

RESULT 4

AAT33536

ID AAT33536 standard; DNA; 15239 BP.

XX AAT33536;

XX 15-FEB-1998 (first entry)

XX BCG deletion region 2 and flanking sequences.

XX BCG delta 2; virulence; avirulence; attenuation; gene deletion;

XX mycobacteria; vaccine; infection; marker; ss.

XX Mycobacterium bovis strain BCG.

XX Key Location/Qualifiers

FT misc\_feature 3382..14071

FT /\*tag= a

FT /note= "BCG delta 1 deletion region"

XX W09G25519-A1.

XX 22-AUG-1996.

XX 15-FEB-1996; 96WO-US01938.

XX 17-FEB-1995; 95US-0390878.

XX (PATH-) PATHOGENESIS CORP.

XX Mahairas GG, Stover CK;

XX WPI; 1996-393419/39.

XX Detecting markers for avirulence in Mycobacterium - used in

XX production of vaccines against bacterial infection, and to detect

XX bacterial infection

XX Example 1; Fig 2; 66pp; English.

XX This DNA sequence comprises Mycobacterium bovis BCG deletion

XX sequence BCGdelta2. A specific genetic deletion of this region

XX results in an avirulence phenotype of the mycobacterium. 2 Other

XX deletion regions (see AAT33535 and AAT33537) have also been detected.

XX Identification involved screening a BCG cosmid library with a

XX radiolabeled probe obtained following DNA subtraction between

XX virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.

XX The deletions provide useful markers for the identification of an

XX avirulent, or a virulent, mycobacterial phenotype. Determination

XX of avirulence requires the detection of the presence or absence of

XX the deletion; the deletions are detected either by detecting the

XX presence or absence of deletion junctions (see AAT33538-46), or by

XX detecting the presence or absence of the sequences contained within

XX the deletion. Deletion polypeptides are used as components of

XX immunological assays and in vaccines.

XX Sequence 15239 BP; 2878 A; 4791 C; 4729 G; 2841 T; 0 other;

Query Match 8.9%; Score 63.2; DB 17; Length 15239;

Best Local Similarity 55.3%; Pred. No. 7.4e-10;

Matches 166; Conservative 0; Mismatches 128; Indels 6; Gaps 2;

QY 418 caactggacacattgtctcattcatattgatgaagcagcttcgaagcgccctcttagcc 477

Db 5253 cagctggcgccattggtcccgctggtcgactggcgacttcgatcgccgcgagcgc 5312

QY 478 ctttccatttccccctggcggtgagtcagcgcttaagactctcgagcatcacgtgggt 537



Db 5313 ctacatgtcaccctcggtgtcagtcagcgcacatcaagtcgttggagcagcaggtcggc 5372  
 Qy 538 cgagtggtggtatcgccaccacccgcccgaagcgaagcgggtgaagtccttgtg 597  
 Db 5373 caggtgctgggtcagggaagccatgctcggcgacgacgcaggtatcccgtgttg 5432  
 Qy 598 caagcagcgcggaaaatggttctgctgcaagcagaaactaaagcgaactatctgg--ac 655  
 Db 5433 cggttgctgcgcgaacacgcttgctgagtcctgagtcggagcgctcgctgaaatgggtggcaac 5492  
 Qy 656 gcttctgtgaa----atcccggttaaccatcgccatcaacgagattcgcctatccacatgg 711  
 Db 5493 gcgtcgctgaaacgcacgcgcggtaccattcggtgtaaacgcgcatccatcgcgacatgg 5552  
 RESULT 5  
 AAA52691/C  
 ID AAA52691 standard; DNA; 636 BP.  
 XX  
 AC AAA52691;  
 XX  
 XX 03-JAN-2001 (first entry)  
 XX  
 XX Escherichia coli yggA gene.  
 XX  
 KW E. coli; yggA gene; amino acid production; excretion protein gene;  
 KW amino acid excretion protein; ds.  
 XX  
 OS Escherichia coli.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..636  
 FT /\*tag= a  
 FT /product= "YggA"  
 XX  
 XX EP1016710-A2.  
 XX  
 PD 05-JUL-2000.  
 XX  
 XX 17-DEC-1999; 99EP-0125263.  
 XX  
 XX 30-DEC-1998; 98RU-0124016.  
 PR 09-MAR-1999; 99RU-0104431.  
 XX  
 XX (AJIN ) AJINOMOTO CO INC.  
 XX  
 XX Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;  
 PI Tokhmakova IL;  
 XX  
 DR WPI; 2000-414802/36.  
 XX  
 XX P-PSDB; AAB01789.  
 XX  
 XX Increased production of L-amino acids by an Escherichia bacterium  
 PT comprises increasing the expression amount of an L-amino acid excretion  
 PT protein -  
 XX  
 PS Disclosure; Page 24; 29pp; English.  
 XX  
 XX The present sequence is the yggA gene (an excretion protein gene) of  
 CC Escherichia coli. The amino acid excretion protein produced from this  
 CC gene is involved in the production of amino acids, and an increase in its  
 CC expression leads to an increased accumulation of amino acids in the cell.  
 CC In this case, an increase in arginine, glutamic acid and lysine is  
 CC achieved if multiple copies of the gene are transfected into a bacterium.  
 CC The bacterium used is E. coli.  
 XX  
 SQ Sequence 636 BP; 120 A; 143 C; 180 G; 193 T; 0 other;

Query Match 8.5%; Score 60.4; DB 21; Length 636;  
 Best Local Similarity 51.1%; Pred. No. 1e-09;  
 Matches 142; Conservative 0; Mismatches 136; Indels 0; Caps 0;

Qy 49 tccacctgttgtcattcggtgttcttctgctgcatgacggcgaacaccataacaggaagcg 108  
 Db 290 TCATATTACTGCTCATTTGCTTTTAAAGCGCCAAACCATACACAGCAGAGGCT 231  
 Qy 109 atgccacccagcgcataataatcagacgacgacgacgacgacgacgacgacgacgacg 168  
 Db 230 ACGCGCCCGCCAGGTGACGCGGCCAGCAACACGCGGCACTGTCATCATATACGCGCTGCCA 171  
 Qy 169 cccaaggtgctggcgatgaacaaaagacgtcagaaaattaaacacacgagagaagaccgca 228  
 Db 170 CCAAAATCCCGCGCAATCAGGACCAATCGCTGATAGCACAAAGTAAGGCAATCATA 111  
 Qy 229 atgagtccttgcgcttaattccttcttataaccaggtacattctgcggtccgattggac 288  
 Db 110 ATGTGGTACTGACGACGTATGCTTATTCATCATAAAAGCATTTTGTGGACCGCGGT 51  
 Qy 289 agtaaaagactggcccccacaaagcagacacctgtaataag 326  
 Db 50 AGGATCATAGCGCCGCCCAAGTGCAAGACCTTGAANGTA 13  
 RESULT 6  
 AAF28535/C  
 ID AAF28535 standard; DNA; 45613 BP.  
 XX  
 AC AAF28535;  
 XX  
 DT 04-APR-2001 (first entry)  
 XX  
 DE Genomic fragment #22.  
 XX  
 KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;  
 KW bronchopulmonary; endocarditis; meningitis; ss.  
 XX  
 OS Moraxella catarrhalis.  
 XX  
 PN WO200078968-A2.  
 XX  
 PD 28-DEC-2000.  
 XX  
 XX 16-JUN-2000; 2000WO-US16649.  
 PF  
 XX 18-JUN-1999; 99US-0140121.  
 PR  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA  
 XX Lagace RE, Patterson C, Berg KL;  
 PI  
 XX WPI; 2001-041427/05.  
 DR  
 XX Genomic library for identifying diagnostic and therapeutic  
 PT compositions, and for identifying virulence factors, regulatory  
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic  
 PT acids -  
 XX  
 PS Claim 1; Page 180-191; 545pp; English.  
 XX  
 XX The present invention relates to a Moraxella catarrhalis genomic library  
 CC comprising a combination of 41 nucleic acid molecules (see  
 CC AAF28514-AAF28554). The library has a number of uses described in the  
 CC specification e.g. is useful for identifying diagnostic and therapeutic  
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
 CC aerobic, gram-negative diplococcus, normally found among the bacterial  
 CC flora of human upper airways. M. catarrhalis is known to cause acute,  
 CC localised infections such as otitis media, sinusitis and bronchopulmonary  
 CC infection and life-threatening, systemic diseases including endocarditis  
 CC and meningitis.  
 XX  
 SQ Sequence 45613 BP; 13249 A; 9346 C; 9947 G; 13071 T; 0 other;

Query Match 6.4%; Score 45.4; DB 22; Length 45613;  
 Best Local Similarity 57.3%; Pred. No. 0.00098;

	Matches	82:	Conservative	0:	Mismatches	61:	Indels	0:	Gaps	0:
Qy	406	atgaacccattcaactggacacttggctccaatcattgatgaagcgagcttcgaaggc	465							
Db	10609	ATGAATACCAACAATCTGCCACTTTTGTTCAGTTATGCACAGCAGCATCTCTAGT	10550							
Qy	466	gcctccttagcccccttccaattccccttcgcggtgagtgcgcggttaaagctctcgag	525							
Db	10549	GCTCGGAAAAGCTTTATATTACGCAACCTCTCTGAAGTAAGCGTATTAAGAATCTTGAA	10490							
Qy	526	catacagctgggtcgagtgttggt	548							
Db	10489	GATCAATTTGGGACGGCGTCTGTT	10467							

RESULT	7	
AA	84317/C	
ID	AA	84317 standard; DNA; 4915 BP.
XX		
XX	AA	84317;
XX		
XX	08-SEP-1999	(first entry)
XX		
XX	Stealth virus	nucleic acid clone, SEQ ID NO: 9.
XX		
XX	Stealth virus;	detection; diagnosis; infection; ss.
XX		
XX	Stealth virus.	
XX		
XX	WO	9934019-A1.
XX		
XX	08-JUL-1999.	
XX		
XX	30-DEC-1998;	98WO-US27744.
XX		
XX	30-DEC-1997;	97US-0001184.
XX		
XX	(MART/)	MARTIN W J.
XX		
XX	Martin WJ;	
XX		
XX	WPI;	1999-405521/34.
XX		
XX	Novel strains of	stealth virus
PT		
PS	Claim 19;	Page 40-42; 95pp; English.
XX		
XX	This sequence	represents a Stealth virus nucleic acid clone. The
CC	invention	relates to a method of detecting and characterising a stealth
CC	virus	by reacting a sample suspected of containing a stealth virus with a
CC	probe	from a known stealth virus and sequencing the resultant isolated
CC	nucleotide.	The method comprises the steps of: (a) isolating DNA or RNA
CC	from a sample	suspected of containing a stealth virus, e.g. a culture of
CC	cells	showing a viral cytopathic effect; (b) testing the reactivity of
CC	the isolated	DNA or RNA with a molecular probe that contains at least 18
CC	or more	contiguous nucleotides identical to sequence previously
CC	identified	from a stealth virus; and, optionally (c) sequencing the
CC	isolated	DNA or RNA molecules that react with the probe. The method is
CC	used	to detect stealth virus in a biological product, food or in the
CC	environment.	The method is also used to evaluate agents for their
CC	inhibitory	or stimulatory effects on stealth virus replication and to
CC	determine	capacity of the virus to recombine with and potentially alter
CC	the nucleic	acid sequences of a cell or bacterium.
XX		
XX	Sequence	4915 BP; 967 A; 1521 G; 1591 G; 828 T; 8 other;

```
Query Match      6.4%; Score 45.2; DB 20; Length 4915;
Best Local Similarity 51.0%; Pred. No. 0.00032;
Matches 107; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
```

Db	1461	CACCTCTGCTGCTGGTTCGCTGACGAGGACTCTGACGGTGCAGCGCATGAACCTCCGCAT	1402
Qy	486	ttccctctcgcggtgaagtcagcggttaaagctctcagcatcacgttggtcgagtggt	545
Db	1401	TACTCCCTCCGCTGTGAGCCAGCGGCTCAAGGCTTTGGAGACTCTTCTGTGTGACGGTCGT	1342
Qy	546	gglatcgcgcaaccacccgagccaaagcaaccgagcggttgaaagtcctttgtgcaagcagc	605
Db	1341	ATTGGTGGCTCAAAAGCCAGTCCGCTGACGAGCACGGGGCGCTCCATCGTTCGGTTTGC	1282
Qy	606	gcggaaaaatggtgtgtgtcgaagcagaaac	635
Db	1281	TCGTCACACCGACATGCTCGACAGGCGAGAC	1252

RESULT	8
AAF58252	
ID	AAF58252 standard; DNA; 936 BP.
XX	
AC	AAF58252;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Oligonucleotide D1835.
XX	
KW	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
XX	
OS	Synthetic.
XX	
PN	WO200107665-A2.
XX	
PD	01-FEB-2001.
XX	
PF	26-JUL-2000; 2000WO-US20476.
XX	
PR	26-JUL-1999; 99US-0145695.
XX	
PR	17-MAR-2000; 2000US-0190259.
XX	
PA	(CLIN-) CLINICAL MICRO SENSORS INC.
XX	
PI	Umek RM;
XX	
PI	WPI; 2001-159728/16.
XX	
DR	
XX	
PT	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface
XX	
XX	
PS	Example 6; Page 127; 159pp; English.
XX	
CC	The present invention relates to a composition comprising two nucleic
CC	acids each containing an electron-transfer group (ETM) having
CC	different redox potentials. The invention is used for electronic
CC	detection of nucleic acids, especially of substitutions (mismatches)
CC	and single-nucleotide polymorphisms, e.g. for genotyping,
CC	monitoring gene expression.
XX	
XX	
SQ	Sequence 936 BP: 4 A: 139 C: 10 G: 7 T: 776 other;

	Query Match	5.3%	Score 37.8;	DB 22;	Length 936;
	Best Local Similarity	5.8%;	Pred. No. 0.034;		
	Matches 18;	Conservative 163;	Mismatches 130;	Indels 0;	Gaps 0;
Qy	184	atgaacaaaagcgtcagaataattaacacacagagaagaaccgcgaatgagtcttcgcgc	243		
	:	: : : :	: : : :	: : : :	: : : :
Db	512	www	www	www	www
Qy	244	ttaattccttgtttaataccacgtagtaccttcgcggtccgatggacagtaaagaactgcc	303		
	:	: : : :	: : : :	: : : :	: : : :
Db	572	www	www	www	www
	:	: : : :	: : : :	: : : :	: : : :
Db	631	www	www	www	www

[illegible]



[illegible]

RESULT	13
AAF58255	
ID	AAF58255 standard; DNA; 938 BP.
XX	
XX	AAF58255;
XX	
XX	24-APR-2001 (first entry)
DT	
XX	
DE	Oligonucleotide D1876.
XX	
XX	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
XX	
XX	Synthetic.
XX	
XX	WO200107665-A2.
PN	
XX	
XX	01-FEB-2001.
PD	
XX	
XX	26-JUL-2000; 2000WO-US20476.
PF	
XX	
XX	26-JUL-1999; 99US-0145695.
PR	
XX	17-MAR-2000; 2000US-0190259.
XX	
XX	(CLIN-) CLINICAL MICRO SENSORS INC.
PA	
XX	
PI	Umek RM;
XX	
DR	WPI; 2001-159728/16.
XX	
PT	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface
XX	
XX	Example 6; Page 127; 159pp; English.
PS	
XX	
CC	The present invention relates to a composition comprising two nucleic
XX	acids each containing an electron-transfer group (ETM) having
CC	different redox potentials. The invention is used for electronic
CC	detection of nucleic acids, especially of substitutions (mismatches)
CC	and single-nucleotide polymorphisms, e.g. for genotyping,
CC	monitoring gene expression.
XX	
XX	Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;
SO	

Query Match 5.3%; Score 37.8; DB 22; Length 938;  
Best Local Similarity 5.8%; Pred. No. 0.034;

	Matches	18; Conservative	163; Mismatches	130; Indels	0; Gaps	0;
Qy	184	atgacacaaaagcgtcagaaaattaaacacacgagaagaccgcaatgagtctcttcgcgc	243			
Db	512	#####	#####	#####	#####	#####
Qy	244	ttaattctctgtttaatcacccagtacattctgcggtccgagtgcagagtaaaagactggcc	303			
Db	572	#####	#####	#####	#####	#####
Qy	304	cccaaaagcagacctgtaatgaagatttccatgatcaccaatcgtgacctatggaagtact	363			
Db	632	#####	#####	#####	#####	#####
Qy	364	taagtaaatgattgggtcttaacatggtttaatatagttcatgaaccatctcaactg	423			
Db	692	#####	#####	#####	#####	#####
Qy	424	gacactttgctctcaatcattgatgaaggcagcttcgaagcgctcctcttagcccttcc	483			
Db	752	#####	#####	#####	#####	#####
Qy	484	attccccctc	494			
Db	812	cccccccccc	822			

RESULT 14

AAN80093/C	
ID	AAN80093 standard; DNA; 2058 BP.
XX	
XX	AAN80093;
XX	
XX	30-MAR-1992 (first entry)
XX	
DE	Sequence of BamHI/Sali fragment of the tfdA gene which encodes 2,4-dichlorophenoxyacetic acid (2,4-D) mono-oxygenase.
DE	
XX	
XX	Herbicide resistance; halogenated aromatic cpd; enzyme;
KW	mono-oxygenase; ss.
KW	
XX	
XX	Alcaligenes eutrophus JMP134.
XX	
XX	
Key	Location/Qualifiers
FFH	748..1608
FFH	/*tag= a
FFH	/note= "DNA SQ claimed"
FFH	
XX	
PN	DE3629890-A.
XX	
XX	
PD	10-MAR-1988.
XX	
XX	
XX	29-AUG-1986; 86DE-3629890.
XX	
XX	29-AUG-1986; 86DE-3629890.
XX	
PR	
XX	(SCHD ) SCHERING AG.
PA	(STRE/) STREBER W.R.
PA	(MACO-) MACQUARIE UNIV.
PA	
XX	
XX	Streber W, Timmis KN, Zenk M;
PI	
XX	
XX	WPI; 1988-071716/11.
DR	P-PSDB; AAP80079.
XX	
XX	
PT	Plasmids and bacteria contg. gene tfdA for 2,4-D-mono:oxygenase -
PT	isolated using new Alcaligenes eutrophus transposon mutants
XX	
XX	
PS	Disclosure; Fig 10; 32pp; German.
XX	
XX	
CC	The tfdA gene product is a 2,4-D decomposing protein. A.eutrophus
CC	JMP 134 contains an 80 kb plasmid, pJP4, which carries the genes
CC	coding for 2,4-D decomposition. Plasmids and bacteria contg. the
CC	

A *tfdA* gene of *P. aeruginosa* plasmid pRO101, which encodes the enzymes for 2,4-D degradation, was cloned into *Agrobacterium tumefaciens* vector pBI121 to form pUCW200. This was used to introduce 2,4-D herbicide-resistance into sweetgum thereby

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:33:09 ; Search time 34.57 Seconds  
(without alignments)  
378,790 Million cell updates/sec

Title: US-09-105-117i-2  
Perfect score: 1091  
Sequence: 1 DPLETMTALRDIIVASGRA.....SHDGINIWAKATDSKTREN 216

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : A\_Geneseq\_0601.\*
- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*
  - 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*
  - 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*
  - 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*
  - 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*
  - 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*
  - 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*
  - 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*
  - 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*
  - 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*
  - 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*
  - 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*
  - 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*
  - 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*
  - 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*
  - 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*
  - 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*
  - 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*
  - 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*
  - 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*
  - 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*
  - 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1091	100.0	216	18 AAW37716	C. glutamicum orf3
2	792	72.6	290	22 AAB76844	Corynebacterium gl
3	225	20.6	348	20 AAV52837	Escherichia coli p
4	205	18.8	329	18 AAW23388	Shaker-like potass
5	202	18.5	329	18 AAW23389	Shaker-like potass
6	195.5	17.9	266	21 AAG35920	Zea mays protein f
7	195.5	17.9	328	21 AAG35919	Zea mays protein f
8	195	17.9	216	21 AAG09266	Arabidopsis thalia
9	195	17.9	216	21 AAG39051	Arabidopsis thalia
10	195	17.9	266	21 AAG39050	Arabidopsis thalia
11	195	17.9	328	21 AAG39049	Arabidopsis thalia

12	184	16.9	312	22 AAB79618	Corynebacterium gl
13	184	16.9	312	22 AAB80067	Corynebacterium gl
14	180.5	16.5	315	22 AAB76845	Zea mays protein f
15	180.5	16.5	316	21 AAG35251	Zea mays protein f
16	180.5	16.5	371	21 AAG35250	Zea mays protein f
17	180.5	16.5	388	21 AAG35249	Sugar biosynthesis
18	175	16.0	333	18 AAW19736	S. erythraea drpp-4
19	175	16.0	333	20 AAW99386	Arabidopsis thalia
20	166.5	15.3	316	21 AAG39500	Arabidopsis thalia
21	166.5	15.3	324	21 AAG39499	Arabidopsis thalia
22	166.5	15.3	365	21 AAG39498	Arabidopsis thalia
23	164	15.0	192	21 AAG09267	Arabidopsis thalia
24	163.5	15.0	357	21 AAG41676	Arabidopsis thalia
25	163.5	15.0	375	21 AAG52050	Arabidopsis thalia
26	163.5	15.0	377	21 AAG52049	Arabidopsis thalia
27	163.5	15.0	386	21 AAG52048	Arabidopsis thalia
28	163.5	15.0	389	21 AAG52008	Arabidopsis thalia
29	163.5	15.0	412	21 AAG41675	Arabidopsis thalia
30	163.5	15.0	419	21 AAG41674	Arabidopsis thalia
31	162.5	14.9	331	20 AAY49837	Human aflatoxin B1
32	162.5	14.9	331	20 AAY49820	Human aflatoxin B1
33	162.5	14.9	331	21 AAB09902	Arabidopsis thalia
34	162	14.8	340	21 AAG38206	Arabidopsis thalia
35	162	14.8	349	21 AAG38205	Arabidopsis thalia
36	162	14.8	351	21 AAG38204	Arabidopsis thalia
37	160.5	14.7	181	21 AAG37809	Arabidopsis thalia
38	160.5	14.7	331	20 AAG3088	Human CPRM3 protei
39	159.5	14.6	330	20 AAG3086	Human CPRM1 protei
40	158	14.5	329	19 AAW1534	Helicobacter polyp
41	156.5	14.3	327	21 AAB09903	Rat aflatoxin B1 a
42	156.5	14.3	338	20 AAW3089	Mouse mCPRM1 allel
43	156.5	14.3	338	20 AAW3090	Mouse mCPRM1 allel
44	154.5	14.2	278	21 AAG36393	Arabidopsis thalia
45	154.5	14.2	286	21 AAG36392	Arabidopsis thalia

ALIGNMENTS

RESULT 1  
AAW37716  
ID AAW37716 standard; Protein; 216 AA.

AC AAW37716;

DT 12-MAR-1998 (first entry)

DE C. glutamicum orf3 protein.

KW LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;

KW Microbial production; amino acid; animal feed additive.

OS Corynebacterium glutamicum.

PN DE19548222-A1.

PD 26-JUN-1997.

PF 22-DEC-1995; 95DE-1048222.

PR 22-DEC-1995; 95DE-1048222.

XX (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.

XX Eggeling L, Sahm H, Vrljic M;

XX WPI; 1997-333867/31.

XX N-PSDB; AAT96816.

XX Increasing microbial production of amino acids, especially lysine -

XX by improving export carrier activity or corresponding gene

XX expression, also new export and regulatory genes from

XX Corynebacterium

PR	09-JUL-1999;	99DE-1032191.
PR	09-JUL-1999;	99DE-1032209.
PR	09-JUL-1999;	99DE-1032212.
PR	09-JUL-1999;	99DE-1032227.
PR	09-JUL-1999;	99DE-1032228.
PR	09-JUL-1999;	99DE-1032229.
PR	09-JUL-1999;	99DE-1032230.
PR	14-JUL-1999;	99DE-1032927.
PR	14-JUL-1999;	99DE-1033005.
PR	14-JUL-1999;	99DE-1033006.
PR	27-AUG-1999;	99DE-1040764.
PR	27-AUG-1999;	99DE-1040765.
PR	27-AUG-1999;	99DE-1040766.
PR	27-AUG-1999;	99DE-1040830.
PR	27-AUG-1999;	99DE-1040831.
PR	27-AUG-1999;	99DE-1040832.
PR	27-AUG-1999;	99DE-1040833.
PR	31-AUG-1999;	99DE-1041378.
PR	31-AUG-1999;	99DE-1041379.
PR	31-AUG-1999;	99DE-1041395.
PR	03-SEP-1999;	99DE-1042077.
PR	03-SEP-1999;	99DE-1042078.
PR	03-SEP-1999;	99DE-1042079.
PR	03-SEP-1999;	99DE-1042088.

27-AUG-1999; 99DE-1040833.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041395.  
PR 03-SEP-1999; 99DE-1042077.  
PR 03-SEP-1999; 99DE-1042078.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042088.  
XX  
XX  
PA (BADI ) BASF AG.  
XX  
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
XX  
XX WPI: 2001-071486/08.  
DR N-PSDB: AAF68077.  
XX  
XX Corynebacterium glutamicum nucleic acids encoding membrane construction  
PT and membrane transport proteins or their portions, useful for typing or  
PT identifying C. glutamicum or related bacteria, and as markers for  
PT transformation :  
PT

PS	Claim 20; Page 1109; 1119pp; English.
XX	
CC	AAC67743 to AAF68080 encode the Corynebacterium glutamicum membrane construction and membrane transport (MCT) proteins given in AAB76510 to AAB76847. The MCT nucleic acids and proteins are useful in the identification of microorganisms which can be used to produce fine chemicals, for modulating fine chemical production in C. glutamicum or related bacteria (e.g. Brevibacterium lactofermentum), the typing or identification of C. glutamicum or related bacteria, as reference points for mapping C. glutamicum genome, and as markers for transformation.
CC	AAF68082 and AAF68082 represent sequencing primers which are used in an example from the present invention.
XX	
SQ	Sequence 290 AA;
Query Match	72.6%; Score 792; DB 22; Length 290;
Best Local Similarity	99.4%; Pred. NO. 4e-72;
Matches 156; Conservative	0; Mismatches 1; Indels 0; Gaps 0
Qy	1 DTPLEETMYALRDIVASGKALYGVISSYPELTAAAEFGCPLLIHOPYSIINRW 60 
Db	131 dtpleetmyalrdivasgkalygvissyypeltacaaefmaeegecpllihopsyiinrw 190 
Qy	61 VEPGGDDGENLLOAANNGLGVATFSPLAQGLLTDTKYLDGTPEGRASQGKSLSEGMLNV 120 
Db	191 veepgddgenllgsaannglgvlatfspaqlldtkydldgtpegrasqgkslsegmlnv 250 
Qy	121 NNTDMVRKLNDIAQRGSQSLAQMALAWVLREGEYGCA 157 
Db	251 nnldmvrklndiaqrqgsilaqmalawvlregeyga 287 
RESULT	3
ID	AAV52837
AD	AAV52837 standard; Protein: 348 AA.

Db	156	ntpieetlealndvwtgkaryigasmshasqfaqaalelqkqhgwaqfsmqdhynliyr	215
Qy	60	WVEEPGGDGENLLQSRANNGLVIAFSPLAAGLLTDKYLDTGPIEGSRASQ-----GKSL	113
Db	216	-----eeeremlplcyqegvavipwsplargltr-----pwgettarlvsdevgknl	263
Qy	114	SEGMLNVNNTDMVRKLNIDIAQERGOSLAQMALAWVLBEQGEYGADVTVSALIGASSVEQL	173
Db	264	ykes-dendaqiaerltgvsceelgatraqvalawllskpg-----laapilgttsreeql	316
Qy	174	DNSLDSLNL	181
Db	317	dellnavd	324
RESULT 4			
AAW23388			
ID	AAW23388 standard; protein; 329 AA.		
XX	AAW23388;		
XX			
DT	02-APR-1998 (first entry)		
XX			
DE	Shaker-like potassium ion channel beta-subunit core region Kv beta 1.		
KW	Shaker-like potassium ion channel; SPC; beta subunit; core region;		
KW	Kv beta 1; N-terminal A and B box; NAB; treatment; cardiac disease;		
KW	tumour; auto immune disease.		
XX			
OS	Mammalia.		
XX			
PN	WO9311112-A2.		
XX			
PD	28-AUG-1997.		
XX			
PF	18-FEB-1997; 97WO-US02292.		
XX			
PR	23-FEB-1996; 96US-0606143.		
XX			

(UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
Li M;  
WPI; 1997-435164/40.  
Polyptide(s) derived from shaker-like potassium ion channel alpha and beta subunits - used to alter potassium ion levels in a cell, e.g. for treating neurological disorders, tumours, metabolic disease and cardiac disease  
Claim 15; Pages 51-52; 106pp; English.  
This polypeptide sequence Kv beta 1 consists of the core region of a beta-subunit of a Shaker-like potassium ion channel (SPC). This core region can bind to a polypeptide consisting of the N-terminal A and B box (NAB) domain and the NAB-S1 (the first transmembrane spanning domain) linking region of the alpha-subunit of SPC. The polypeptides or the nucleic acid encoding them can be introduced into the cytoplasm of a cell to modulate the flow of potassium ions through a cytoplasmic cell membrane. Potassium ion channels regulate the action potentials, card pacemaking and neurotransmitter release in excitable tissues. In non-excitable tissues they play important roles in hormone secretion, cell proliferation, cell volume regulation and lymphocyte differentiation. Molecules which bind to the alpha or beta-subunit polypeptides can be detected by contacting the polypeptides, with a putative NAB and NAB-S1 linking region of an alpha-subunit or with a putative core region of a beta-subunit respectively, and determining whether or not binding occurs. These polypeptides and the encoding nucleic acids may be useful in the treatment of a huge variety of disorders, e.g. neurological disorders, tumours, metabolic diseases, cardiac disease and autoimmune disease.

CC putative core region of a beta-subunit respectively, and determining  
CC whether or not binding occurs. These polypeptides and the encoding  
CC nucleic acids may be useful in the treatment of a huge variety of  
CC disorders, e.g. neurological disorders, tumours, metabolic diseases,  
CC cardiac disease and autoimmune disease.

This polypeptide sequence Kv beta 2 consists of the core region of a beta-subunit of a Shaker-like potassium ion channel (SPC). This core region can bind to a polypeptide consisting of the N-terminal A and B box (NAB) domain and the NAB-S1 (the first transmembrane spanning domain), linking region of the alpha-subunit of SPC. The polypeptides or the nucleic acid encoding them can be introduced into the cytoplasm of a cell to modulate the flow of potassium ions through a cytoplasmic cell membrane. Potassium ion channels regulate the action potentials, cardiac pacemaking and neurotransmitter release in excitable tissues. In non-excitable tissues they play important roles in hormone secretion, cell proliferation, cell volume regulation and lymphocyte differentiation. Molecules which bind to the alpha or beta-subunit polypeptides can be detected by contacting the polypeptides with a

PR	01-APR-1999	9905-0127462.
PR	06-APR-1999	9905-0128234.
PR	08-APR-1999	9905-0128714.
PR	16-APR-1999	9905-0129845.
PR	19-APR-1999	9905-0130077.
PR	21-APR-1999	9905-0130449.
PR	23-APR-1999	9905-0130510.
PR	23-APR-1999	9905-0130891.
PR	28-APR-1999	9905-0131449.
PR	30-APR-1999	9905-0132048.
PR	30-APR-1999	9905-0132407.
PR	04-MAY-1999	9905-0132484.
PR	05-MAY-1999	9905-0132485.
PR	06-MAY-1999	9905-0132486.

PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 18-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140895.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159394.  
PR 13-OCT-1999; 99US-0159395.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.



PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 28-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149426.  
 PR 18-AUG-1999; 99US-0149472.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.

PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 22-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 17.9%; Score 195.5; DB 21; Length 328;  
 Best Local Similarity 30.1%; Pred. No. 1.4e-11;  
 Matches 65; Conservative 39; Mismatches 89; Indels 23; Gaps 8;

QY 2 TPLEETWALRDIVASGKALYGVISYGPETA---AEPMAEEGCPLLIHQPSYSIIN 58  
 Db 128 tpveetvrainwldqgwfytsewsaqgteawavanrldivg--piveqpeynlfs 185  
 QY 59 RWVEEPGDDGENLLOSAANGLGVIAFSPLAOGLLTDKYLDG-IPEGSR--ASQGKSLSE 115  
 Db 186 rhkve-----seflpystygtgittwspiasgvltyktnipadsrfaldnyknlan 240  
 QY 116 GMLNVNIDMVRKLNIAQERGQSLAQMALAWVLREGQEGYADTVTSALIGASSVEQLDN 175  
 Db 241 rslvddtlrkvnlgkplaaelgvslqslawcasn-----pnvssvitgatkessqive 294  
 QY 176 SLDSLNLEFSDAELEAIDEISHDAGINIWAKATDS 211  
 Db 295 nmkaldivliltpev--vdri--eavvgskpkrtes 326

RESULT 8

AAG09266  
 ID AAG09266 standard; Protein: 216 AA.

AC AAG09266;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 7130.

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-APR-1999; 99US-0144335.  
PR 21-APR-1999; 99US-0144352.  
PR 23-APR-1999; 99US-0144632.  
PR 28-APR-1999; 99US-0144884.  
PR 30-APR-1999; 99US-0144814.  
PR 30-APR-1999; 99US-0145086.  
PR 04-MAY-1999; 99US-0145088.  
PR 05-MAY-1999; 99US-0145085.  
PR 06-MAY-1999; 99US-0145087.  
PR 06-MAY-1999; 99US-0145089.  
PR 07-MAY-1999; 99US-0145192.  
PR 11-MAY-1999; 99US-0145145.  
PR 14-MAY-1999; 99US-0145218.  
PR 14-MAY-1999; 99US-0145224.  
PR 14-MAY-1999; 99US-0145276.  
PR 14-MAY-1999; 99US-0145913.  
PR 14-MAY-1999; 99US-0145918.  
PR 18-MAY-1999; 99US-0145919.  
PR 19-MAY-1999; 99US-0145951.  
PR 20-MAY-1999; 99US-0146386.  
PR 21-MAY-1999; 99US-0146388.  
PR 24-MAY-1999; 99US-0146389.  
PR 25-MAY-1999; 99US-0147038.  
PR 27-MAY-1999; 99US-0147204.  
PR 01-JUN-1999; 99US-0147302.  
PR 03-JUN-1999; 99US-0147192.  
PR 04-JUN-1999; 99US-0147260.  
PR 07-JUN-1999; 99US-0147303.  
PR 08-JUN-1999; 99US-0147416.  
PR 10-JUN-1999; 99US-0147493.  
PR 10-JUN-1999; 99US-0147935.  
PR 14-JUN-1999; 99US-0148171.  
PR 16-JUN-1999; 99US-0148319.  
PR 16-JUN-1999; 99US-0148341.  
PR 17-JUN-1999; 99US-0148565.  
PR 18-JUN-1999; 99US-0148684.  
PR 18-JUN-1999; 99US-0149368.  
PR 18-JUN-1999; 99US-0149175.  
PR 18-JUN-1999; 99US-0149426.  
PR 18-JUN-1999; 99US-0149722.  
PR 18-JUN-1999; 99US-0149723.  
PR 18-JUN-1999; 99US-0149929.  
PR 18-JUN-1999; 99US-0149902.  
PR 18-JUN-1999; 99US-0149930.  
PR 18-JUN-1999; 99US-0150566.  
PR 18-JUN-1999; 99US-0150884.  
PR 18-JUN-1999; 99US-0151065.  
PR 18-JUN-1999; 99US-0151066.  
PR 21-JUN-1999; 99US-0151080.  
PR 22-JUN-1999; 99US-0151303.  
PR 23-JUN-1999; 99US-0151438.  
PR 23-JUN-1999; 99US-0151930.  
PR 24-JUN-1999; 99US-0152363.  
PR 28-JUN-1999; 99US-0153070.  
PR 29-JUN-1999; 99US-0153758.  
PR 30-JUN-1999; 99US-0154018.  
PR 01-JUL-1999; 99US-0154039.  
PR 01-JUL-1999; 99US-0154779.  
PR 02-JUL-1999; 99US-0155139.  
PR 06-JUL-1999; 99US-0155486.  
PR 08-JUL-1999; 99US-0155659.  
PR 09-JUL-1999; 99US-0156458.  
PR 12-JUL-1999; 99US-0156596.  
PR 13-JUL-1999; 99US-0157117.  
PR 14-JUL-1999; 99US-0157753.  
PR 15-JUL-1999; 99US-0157865.  
PR 16-JUL-1999; 99US-0158029.  
PR 16-JUL-1999; 99US-0158232.  
PR 19-JUL-1999; 99US-0158369.  
PR 19-JUL-1999; 99US-0159293.  
PR 19-JUL-1999; 99US-0159294.  
PR 19-JUL-1999; 99US-0159295.  
PR 19-JUL-1999; 99US-0159329.



PR	14-OCT-1999;	99US-0159330.	PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	PR	01-APR-1999;	99US-0127462.	
PR	14-OCT-1999;	99US-0159638.	PR	06-APR-1999;	99US-0128234.	
PR	18-OCT-1999;	99US-0159584.	PR	08-APR-1999;	99US-0128714.	
PR	21-OCT-1999;	99US-0160741.	PR	16-APR-1999;	99US-0129845.	
PR	21-OCT-1999;	99US-0160767.	PR	19-APR-1999;	99US-0130077.	
PR	21-OCT-1999;	99US-0160768.	PR	21-APR-1999;	99US-0130449.	
PR	21-OCT-1999;	99US-0160770.	PR	23-APR-1999;	99US-0130510.	
PR	21-OCT-1999;	99US-0160814.	PR	23-APR-1999;	99US-0130891.	
PR	21-OCT-1999;	99US-0160815.	PR	28-APR-1999;	99US-0131449.	
PR	22-OCT-1999;	99US-0160980.	PR	30-APR-1999;	99US-0132048.	
PR	22-OCT-1999;	99US-0160981.	PR	30-APR-1999;	99US-0132407.	
PR	22-OCT-1999;	99US-0160989.	PR	04-MAY-1999;	99US-0132484.	
PR	25-OCT-1999;	99US-0161404.	PR	05-MAY-1999;	99US-0132485.	
PR	25-OCT-1999;	99US-0161405.	PR	06-MAY-1999;	99US-0132486.	
PR	25-OCT-1999;	99US-0161406.	PR	06-MAY-1999;	99US-0132487.	
PR	26-OCT-1999;	99US-0161359.	PR	07-MAY-1999;	99US-0132863.	
PR	26-OCT-1999;	99US-0161360.	PR	11-MAY-1999;	99US-0134256.	
PR	26-OCT-1999;	99US-0161361.	PR	14-MAY-1999;	99US-0134218.	
PR	28-OCT-1999;	99US-0161920.	PR	14-MAY-1999;	99US-0134219.	
PR	28-OCT-1999;	99US-0161992.	PR	14-MAY-1999;	99US-0134221.	
PR	28-OCT-1999;	99US-0161993.	PR	14-MAY-1999;	99US-0134370.	
PR	29-OCT-1999;	99US-0162142.	PR	18-MAY-1999;	99US-0134768.	
PR	29-OCT-1999;	99US-0162142.	PR	19-MAY-1999;	99US-0134768.	
Query Match					99US-0134941.	
Best Local Similarity					99US-0135124.	
Matches 61; Conservative 41; Mismatches 72; Indels 36; Gaps 9;					99US-0135353.	
QY	2	TPLEETMYALRDIVASGKALYGVSSGPELTAAEFMAEEGCP---	LIHQPSYSIIN 58	PR	21-MAY-1999;	99US-0135629.
Db	16	tlpleetramnyvldkqwfvgwtsewsaqqitea--wgaadrldlvgpaveqpeymnfa 73		PR	24-MAY-1999;	99US-0136021.
QY	59	RWBEPEGDGCENLQSAANGLGVIAFSPLAQGLLTDKYLIDG-IPEGSRASQGRSLSEGM 117		PR	25-MAY-1999;	99US-0136392.
Db	74	rhkve-----teflplynthgigttsplasgvtlkyngkaipsdsrfa-----l 120		PR	27-MAY-1999;	99US-0136782.
QY	118	LNWNVI-----DMVRK---LNDIAERQGSQAQMALAWLVRQEGYGDVTVTSALIGA 167		PR	28-MAY-1999;	99US-0137222.
Db	121	enyknlansrldvdvkrksglkipadclgvtlaqlaiawcasn-----pnvssvitga 174		PR	01-JUN-1999;	99US-0137528.
QY	168	SSVEQLDNLSDLNNLE-FSDAELEAIDEI 196		PR	03-JUN-1999;	99US-0137502.
Db	175	tresqigemkavdviplltpivdkieqv 204		PR	04-JUN-1999;	99US-0137724.
RESULT 9					99US-0137724.	
XX	AAG39051	standard; Protein; 216 AA.		PR	07-JUN-1999;	99US-0138094.
XX	AC	AAG39051;		PR	08-JUN-1999;	99US-0138540.
XX	XX			PR	10-JUN-1999;	99US-0138847.
XX	DT	18-OCT-2000 (first entry)		PR	10-JUN-1999;	99US-0139119.
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 48265.		PR	14-JUN-1999;	99US-0139452.
XX	XX	Protein identification; signal transduction pathway; metabolic pathway;		PR	16-JUN-1999;	99US-0139453.
XX	KW	hybridisation assay; genetic mapping; gene expression control; promoter;		PR	17-JUN-1999;	99US-0139457.
XX	KW	termination sequence.		PR	18-JUN-1999;	99US-0139458.
XX	OS	Arabidopsis thaliana.		PR	18-JUN-1999;	99US-0139459.
XX	XX	EP1033405-A2.		PR	18-JUN-1999;	99US-0139461.
XX	PD	06-SEP-2000.		PR	18-JUN-1999;	99US-0139462.
XX	PF	25-FEB-2000; 2000EP-0301439.		PR	18-JUN-1999;	99US-0139463.
XX	XX	25-FEB-1999; 99US-0121825.		PR	18-JUN-1999;	99US-0139750.
XX	PR	05-MAR-1999; 99US-0123180.		PR	18-JUN-1999;	99US-0139763.
XX	PR	09-MAR-1999; 99US-0123548.		PR	21-JUN-1999;	99US-0139817.
XX	PR	23-MAR-1999; 99US-0125788.		PR	22-JUN-1999;	99US-0139899.
XX	PR	23-MAR-1999; 99US-0126264.		PR	23-JUN-1999;	99US-0140353.
XX	PR			PR	24-JUN-1999;	99US-0140354.
XX	PR			PR	28-JUN-1999;	99US-0140695.
XX	PR			PR	29-JUN-1999;	99US-0140823.
XX	PR			PR	30-JUN-1999;	99US-0140991.
XX	PR			PR	01-JUL-1999;	99US-0141287.
XX	PR			PR	01-JUL-1999;	99US-0141842.
XX	PR			PR	02-JUL-1999;	99US-0142055.
XX	PR			PR	06-JUL-1999;	99US-0142390.
XX	PR			PR	08-JUL-1999;	99US-0142803.
XX	PR			PR	09-JUL-1999;	99US-0142920.
XX	PR			PR	12-JUL-1999;	99US-0142977.
XX	PR			PR	13-JUL-1999;	99US-0143542.
XX	PR			PR	14-JUL-1999;	99US-0143624.
XX	PR			PR	15-JUL-1999;	99US-0144005.
XX	PR			PR	16-JUL-1999;	99US-0144085.
XX	PR			PR	16-JUL-1999;	99US-0144086.



PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 21-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144864.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
PR 29-OCT-1999; 99US-0162142.

Query Match 17.9%; Score 195; DB 21; Length 266;

Best Local Similarity 29.0%; Pred. No. 1.2e-11;

Matches 61; Conservative 41; Mismatches 72; Indels 36; Gaps 9;

QY 2 TPLEETMYALRDIVASGKALYGISSYGPGLTAEAEFMAEEGCPL---LIHOPSYSIIN 58  
Db 66 tpleetvramnyvidkgwafygtsewsaqqitea--wgaadrldvgpiveqeynmfa 123  
QY 59 RWVEEPDGDGNLLQSAANNGLVAFSPLAQGLLTDKYLDG-IPEGRASQCKSLSEGM 117  
Db 124 rhkve-----teflplythnhglttwsplagvltgkynkgaipsdsrfa-----l 170  
QY 118 LNVNNT-----DMVRK---LNDIAQERGSLAQMALAWLREQGEYGADTVTSALIGA 167  
Db 171 enyknlanrslvddvirkvsgkpiadelgvtlaqlaiawcasn-----pnvssvitga 224  
QY 168 SSVEQLDNLSENLLE-FSDAELENAIDEI 196  
Db 225 tresqigenmkavdvipllitpivldkieqv 254

RESULT 11

AAG39049

ID AAG39049 standard; Protein; 328 AA.

XX AC AAG39049;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 48263.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.  
PD 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 17.9%; Score 195; DB 21; Length 328;

Best Local Similarity 29.0%; Pred. No. 1.6e-11;

Matches 61; Conservative 41; Mismatches 72; Indels 36; Gaps 9;

Qy 2 TPLEETMYALRDIVASGKALYVIGSYGPELTAEAEFMAEEGCP---LIHQPSYSIIN 58

Db 128 tpieetvramnyvidkgwafywtsewsaqitea--wgaadridlvgplveqpeymnfa 185

Qy 59 RWVEEPGDGGENLLQSANNGLVIAFSPLAQGLLTDKYLDG-IPEGSRASQKSLSEGM 117

Db 186 rhkve-----teflplytnhgigtwtspiasgvltkynkgaiptsrfa-----l 232

Qy 118 LNVNNI-----DMVRK---LNDIAOERGQSLAQMALAWVLREQGEYGDVTVTSALIGA 167

Db 233 enyknlanrslvddvirkvsglkipiadelgvltlaqlaiawcasn-----pnvssvitga 286

Qy 168 SSVEQLDNLSDLNLE-FSDAELEAIDEI 196

Db 287 tresqiqenmkavdviplltpivldkieqv 316

RESULT 12

AAB79618

ID AAB79618 standard; Protein; 312 AA.

XX AAB79618;

XX AC

XX 30-APR-2001 (first entry)

XX Corynebacterium glutamicum SMP protein sequence SEQ ID NO:752.

DE DE

XX Corynebacterium glutamicum; carbon metabolism and energy production;

XX KW

KW SMP protein; sugar metabolism and oxidative phosphorylation protein;

99US-0154779.

fine chemical production; organic acid; proteinogenic amino acid; nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside; nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol; carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme; diagnosis; *Corynebacterium diphtheriae*; evolutionary study.

*Corynebacterium glutamicum*.

WO200100844-A2.

04-JAN-2001.

23-JUN-2000; 2000WO-IB00943.

25-JUN-1999: 99US-0141031

08-JUL-1999; 99DE-1031412;

08-JUL-1999; 99DE-1031413.

08-JUL-1999; 99DE-1031419.

08-JUL-1999; 99DE-1031420.

08-JUL-1999; 99DE-1031424.  
08-JUL-1999; 99DE-1031438

08-JUL-1999; 99DE-1031428.  
08-JUL-1999; 99DE-1031431

08-JUL-1999; 99DE-1031431.  
08-JUL-1999; 99DE-1031433.

08-JUL-1999; 99DE-1031434;

08-JUL-1999; 99DE-1031510.

08-JUL-1999; 99DE-1031562.

08-JUL-1999; 99DE-1031634.

09-JUL-1999; 99DE-1032180.  
09-JUL-1999; 99DE-1033337

09-JUL-1999; 99DE-1032227.  
09-JUL-1999. 99DE-1032230

09-JUL-1999: 99DE-1032230.  
09-JUL-1999: 99US-0143208.

14-JUL-1999; 99DE-1032924;

14-JUL-1999; 99DE-1032973.

14-JUL-1999; 99DE-1033005.

27-AUG-1999; 99DE-1040765.

31-AUG-1999; 99US-0151572.  
03-SEP-1999; 99DE-104307E

03-SEP-1999; 99DE-1042076.  
03-SEP-1999; 99DE-1042079

03-SEP-1999; 99DE-1042079.  
03-SEP-1999; 99DE-1042086.

03-SEP-1999; 99DE-1042087;

03-SEP-1999; 99DE-1042088.

03-SEP-1999; 99DE-1042095.

03-SEP-1999; 99DE-1042123.

03-SEP-1999; 99DE-1042125.

(BADT ) BASE AG

(BADI) BASE AG.

Pompejus M, Kroeger B, Sch

[illegible]

WPI; 2001-061975

N-PSDB; AAF71735.

Nov. 1900; 1000

New Isolated Cory  
metabolism and ov

metabolism and oxidative phosphorylation protein for production or modulation of production of fine chemicals e.g. amino acids, carbohydrates or enzymes -

Claim 20; Page 1208-1209; 1246pp; English.

AAF71360 to AAF71750 encode the *Corynebacterium glutamicum* sugar metabolism and oxidative phosphorylation (SMP) proteins given in AAB79243 to AAB 79633 which are involved in carbon metabolism and energy production. The *C. glutamicum* SMP gene can be used in vectors (II) for expression in host cells and production or modulation of production of fine chemicals, such as, an organic acid, a proteinogenic or nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins (III) encoded by them are used for diagnosing the presence or activity of *Corynebacterium diphtheriae* in a subject. (I), (II), (III) or host cells containing them are used to map genomes of organisms related to *C. glutamicum*, identify and localise *C. glutamicum* sequences of interest.



```

XX  Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX  WPI; 2001-071486/08.
XX  N-PSDB; AAF68078.
XX  Corynebacterium glutamicum nucleic acids encoding membrane construction
XX  and membrane transport proteins or their portions, useful for typing or
XX  identifying C. glutamicum or related bacteria, and as markers for
XX  transformation
XX  Claim 20; Page 1111-1112; 1119pp; English.
XX  AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
XX  construction and membrane transport (MCT) proteins given in AAF6510 to
XX  AAF6847. The MCT nucleic acids and proteins are useful in the
XX  identification of microorganisms which can be used to produce fine
XX  chemicals, for modulating fine chemical production in C. glutamicum or
XX  related bacteria (e.g. Brevibacterium lactofermentum), the typing or
XX  identification of C. glutamicum or related bacteria, as reference points
XX  for mapping C. glutamicum genome, and as markers for transformation.
XX  AAF68082 and AAF68082 represent sequencing primers which are used in an
XX  example from the present invention.
XX  Sequence 315 AA;
XX  Query Match 16.5%; Score 180.5; DB 22; Length 315;
XX  Best Local Similarity 29.4%; Pred. NO. 4.4e-10;
XX  Matches 58; Conservative 38; Mismatches 80; Indels 21; Gaps 6;
XX  QY 4 LEETVMDRDIVASGKALYGVISGPELTAAAEFMAEGCPLLIHQPSYSINRWEE 63
XX  Db 133 ideqvaiaidlaqgkikhlaishysaerlaeffeksvgtpaqpvalqphynlvsi- 188
XX  QY 64 PGDDGNLLQSANNGLVAFSPLAQGLTDKYL--DGIPEGRASQ-GKSLSEGMLNV 120
XX  Db 189 -kdyeevnpqlaekghvavfpyfalaagiltgkytskediskaragqldryasdeafv 247
XX  QY 121 NNIDMYRKLNDIAQERGQSLAQMALAWLREOGEYGAADTVTSALIGASSVEQLDNLDSL 180
XX  Db 248 -----vteiravadelgsvpttvalawiv-----ahgtaplasvskveql-kdlmav 294
XX  QY 181 NNLEFSDAELEAIDEIS 197
XX  Db 295 kdvelsaeglarldkvs 311
XX  RESULT 15
XX  AAG35251
XX  ID AAG35251 standard; Protein; 316 AA.
XX  AC AAG35251;
XX  DT 18-OCT-2000 (first entry)
XX  DE Zea mays protein fragment SEQ ID NO: 43035.
XX  KW Protein identification; signal transduction pathway; metabolic pathway;
XX  KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX  KW termination sequence; corn.
XX  OS Zea mays subsp. mays.
XX  PN EP1033405-A2.
XX  XX EP1033405-A2.
XX  PD 06-SEP-2000.
XX  XX 25-FEB-2000; 2000EP-0301439.
XX  PF 25-FEB-1999; 99US-0121825.
XX  PR 05-MAR-1999; 99US-0123180.
XX  PR 09-MAR-1999; 99US-0123548.
XX  PR 23-MAR-1999; 99US-0125788.
XX  PR 25-MAR-1999; 99US-0126264.
XX  PR 29-MAR-1999; 99US-0126785.
XX  PR 01-APR-1999; 99US-0127462.
XX  PR 06-APR-1999; 99US-0128234.
XX  PR 08-APR-1999; 99US-0128714.
XX  PR 16-APR-1999; 99US-0129845.
XX  PR 19-APR-1999; 99US-0130077.
XX  PR 21-APR-1999; 99US-0130449.
XX  PR 23-APR-1999; 99US-0130510.
XX  PR 23-APR-1999; 99US-0130891.
XX  PR 28-APR-1999; 99US-0131449.
XX  PR 30-APR-1999; 99US-0132048.
XX  PR 30-APR-1999; 99US-0132407.
XX  PR 04-MAY-1999; 99US-0132484.
XX  PR 05-MAY-1999; 99US-0132485.
XX  PR 06-MAY-1999; 99US-0132486.
XX  PR 06-MAY-1999; 99US-0132487.
XX  PR 07-MAY-1999; 99US-0132863.
XX  PR 11-MAY-1999; 99US-0134256.
XX  PR 14-MAY-1999; 99US-0134218.
XX  PR 14-MAY-1999; 99US-0134219.
XX  PR 14-MAY-1999; 99US-0134221.
XX  PR 14-MAY-1999; 99US-0134370.
XX  PR 18-MAY-1999; 99US-0134768.
XX  PR 19-MAY-1999; 99US-0134941.
XX  PR 20-MAY-1999; 99US-0135124.
XX  PR 21-MAY-1999; 99US-0135353.
XX  PR 21-MAY-1999; 99US-0135629.
XX  PR 24-MAY-1999; 99US-0136021.
XX  PR 27-MAY-1999; 99US-0136392.
XX  PR 28-MAY-1999; 99US-0136782.
XX  PR 01-JUN-1999; 99US-0137222.
XX  PR 03-JUN-1999; 99US-0137528.
XX  PR 04-JUN-1999; 99US-0137502.
XX  PR 07-JUN-1999; 99US-0137724.
XX  PR 08-JUN-1999; 99US-0138094.
XX  PR 10-JUN-1999; 99US-0138540.
XX  PR 10-JUN-1999; 99US-0138847.
XX  PR 14-JUN-1999; 99US-0139119.
XX  PR 16-JUN-1999; 99US-0139452.
XX  PR 16-JUN-1999; 99US-0139453.
XX  PR 17-JUN-1999; 99US-0139492.
XX  PR 18-JUN-1999; 99US-0139454.
XX  PR 18-JUN-1999; 99US-0139455.
XX  PR 18-JUN-1999; 99US-0139456.
XX  PR 18-JUN-1999; 99US-0139457.
XX  PR 18-JUN-1999; 99US-0139458.
XX  PR 18-JUN-1999; 99US-0139459.
XX  PR 18-JUN-1999; 99US-0139460.
XX  PR 18-JUN-1999; 99US-0139461.
XX  PR 18-JUN-1999; 99US-0139462.
XX  PR 18-JUN-1999; 99US-0139463.
XX  PR 18-JUN-1999; 99US-0139750.
XX  PR 18-JUN-1999; 99US-0139763.
XX  PR 21-JUN-1999; 99US-0139817.
XX  PR 23-JUN-1999; 99US-0139899.
XX  PR 23-JUN-1999; 99US-0140353.
XX  PR 23-JUN-1999; 99US-0140354.
XX  PR 24-JUN-1999; 99US-0140695.
XX  PR 28-JUN-1999; 99US-0140823.
XX  PR 29-JUN-1999; 99US-0140991.
XX  PR 30-JUN-1999; 99US-0141287.
XX  PR 01-JUL-1999; 99US-0141842.
XX  PR 02-JUL-1999; 99US-0142154.
XX  PR 06-JUL-1999; 99US-0142055.
XX  PR 06-JUL-1999; 99US-0142390.
XX  PR 08-JUL-1999; 99US-0142803.
XX  PR 09-JUL-1999; 99US-0142920.
XX  PR 12-JUL-1999; 99US-0142977.
XX  PR 13-JUL-1999; 99US-0143542.
XX  PR 14-JUL-1999; 99US-0143624.
XX  PR 15-JUL-1999; 99US-0144005.

```



```

PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 17-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0149368.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

```

```

PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.5%; Score 180.5; DB 21; Length 316;
Best Local Similarity 26.3%; Pred. No. 4.4e-10;
Matches 52; Conservative 45; Mismatches 76; Indels 25; Gaps 6;

QY 3 PLEETMY--ALRDIVASGKALYGVIGISSYGPFLTAFAAEFMAEEGCPLLIHQPSYSIINRW 60
Db 123 plqeralwdglvemyekglvsavgsnygpkqlkikhsylasrgvplssaqvqfslism- 181
QY 61 VEEPGDDGENLQSAANNGLVIAFSPLAOGLLTDKY-LDGIPEGSRASOGKSLSEGMN 119
Db 182 ----gddqmelksvcdslgltayspglgmltgkygasnlpkgprsvifrqilpl-- 235
QY 120 VNNIDMYVKLNDIAQERGQSLAQMALAWLREOGEYGADTVTSALIGASSVEQOLDNSLDS 179
Db 236 ---epllscrlrlaekkgkmpqvainwcmckg-----tvpipgyktvrhwqenlga 284
QY 180 ----LNNLEFSDAELEAI 193
Db 285 lgwrlssaeseleesaaam 302

```

Search completed: September 13, 2001, 17:51:44  
Job time: 1115 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:51:50 ; Search time 37.95 Seconds  
(without alignments)  
753.040 Million cell updates/sec

Title: US-09-105-117i-2

Perfect score: 1091

Sequence: 1 DTPLEETWYALRDIVASGKA.....SHDAGINIWAKATDSKTREN 216

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTEMBL16:\*  
2: sp-archaea:\*  
3: sp-bacteria:\*  
4: sp-fungi:\*  
5: sp-invertebrate:\*  
6: sp-mammal:\*  
7: sp-mhc:\*  
8: sp-organelle:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-unclassified:\*  
13: sp-vertebrate:\*  
14: sp-virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1054	96.6	209	2 P94634	P94634 corynebacte
2	529.5	48.5	346	2 Q46851	Q46851 escherichia
3	529	48.5	348	2 Q9EWJ0	Q9EWJ0 streptomyc
4	449	41.2	319	2 Q9X7S9	Q9X7S9 streptomyc
5	438.5	40.2	334	2 Q9ZB20	Q9ZB20 lactococcus
6	283.5	26.0	315	2 Q9RS10	Q9RS10 deinococcus
7	267	24.5	319	2 Q9WYE9	Q9WYE9 thermotoga
8	257	23.6	336	1 Q34185	Q34185 halobacteri
9	238	21.8	336	1 Q9H0U7	Q9H0U7 halobacteri
10	237.5	21.8	387	2 Q9PGD5	Q9PGD5 xylella fas
11	227.5	20.9	344	3 Q59826	Q59826 schizosacch
12	213	19.5	354	2 Q44328	Q44328 agrobacteri
13	211	19.3	345	2 Q9HWY1	Q9HWY1 pseudomonas
14	210.5	19.3	330	2 Q9FZ25	Q9FZ25 streptomyc
15	210.5	19.3	336	2 Q9LIS0	Q9LIS0 streptomyc
16	209	19.2	307	2 Q9RJCA	Q9RJCA streptomyc
17	209	19.2	330	10 Q82064	Q82064 solanum tub
18	209	19.2	367	6 Q27955	Q27955 bos taurus
19	209	19.2	367	6 Q9NOV9	Q9NOV9 oryctolagus

20	208	19.1	367	11	Q64284	Q64284 mus musculus
21	206.5	18.9	316	2	Q9HXX2	Q9HXX2 pseudomonas
22	206	18.9	367	4	Q13303	Q13303 homo sapien
23	205	18.8	401	11	Q63277	Q63277 mus musculus
24	205	18.8	408	6	Q28528	Q28528 mustela put
25	205	18.8	419	4	Q14722	Q14722 homo sapien
26	205	18.8	419	6	Q9XT31	Q9XT31 oryctolagus
27	204.5	18.7	401	13	Q9PWR1	Q9PWR1 gallus gall
28	202	18.5	401	13	Q9PTM4	Q9PTM4 xenopus lae
29	199	18.2	367	13	Q9PTM5	Q9PTM5 xenopus lae
30	197	18.1	305	1	Q9HLE1	Q9HLE1 thermoplas
31	197	18.1	352	2	Q9KU57	Q9KU57 vibrio chol
32	196	18.0	324	1	Q9H9H9	Q9H9H9 halobacteri
33	195	17.9	328	10	Q49847	Q49847 egeria dens
34	195	17.9	328	10	Q23016	Q23016 arabidopsis
35	194.5	17.8	316	2	Q05408	Q05408 bacillus su
36	193	17.7	327	2	Q9RS66	Q9RS66 deinococcus
37	192	17.6	328	10	Q39151	Q39151 arabidopsis
38	191	17.5	314	2	Q55493	Q55493 synechocyst
39	189	17.3	331	2	Q91009	Q91009 pseudomonas
40	186.5	17.1	329	2	Q9PCQ4	Q9PCQ4 xylella fas
41	184.5	16.9	330	2	Q9EWF0	Q9EWF0 streptomyc
42	183.5	16.8	362	2	Q9PCQ5	Q9PCQ5 xylella fas
43	181	16.6	343	1	Q9HI90	Q9HI90 thermoplas
44	180	16.5	274	2	Q9X265	Q9X265 thermotoga
45	178.5	16.4	338	10	Q22627	Q22627 helianthus

#### ALIGNMENTS

RESULT 1

P94634 ID P94634 PRELIMINARY; PRT; 209 AA.  
AC P94634;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997\* (TrEMBLrel. 03, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE LYSE AND LYSO GENES.  
OS Corynebacterium glutamicum (Brevibacterium flavum).  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
OC Corynebacterium.  
OX NCBI\_taxid=1718;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R127;  
RX MEDLINE=97126810; PubMed=8971704;  
RA Vrljic M., Sahm H., Eggeling L.;  
RT "A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum.";  
RL Mol. Microbiol. 22:815-826(1996).  
DR EMBL: X96471; CAA65325.1;  
SQ SEQUENCE 209 AA; 22450 MW; 06549D44F0BC0100 CRC64;

Query Match 96.6%; Score 1054; DB 2; Length 209;  
Best Local Similarity 100.0%; Pred. No. 1.2e-75;  
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	8	MYALRDIVASGKALYGVIGISSYSGELTAAEFMAEECCPLLHOPSYSIINRWVEEPDGD	67
DB	1	MYALRDIVASGKALYGVIGISSYSGELTAAEFMAEECCPLLHOPSYSIINRWVEEPDGD	60
QY	68	GENLQSAANGLGVTAFAFSLAQLGTLTKYLDGIPESGRASQKSLSEGLMNNVNDMVR	127
DB	61	GENLQSAANGLGVTAFAFSLAQLGTLTKYLDGIPESGRASQKSLSEGLMNNVNDMVR	120
QY	128	KLNDIAQERGQSQAOMALAWLWREQGEYGDVTYSALIGASSVEQLDNLSDLNLEFSD	187
DB	121	KLNDIAQERGQSQAOMALAWLWREQGEYGDVTYSALIGASSVEQLDNLSDLNLEFSD	180
QY	188	ALEAIDEISHDAGINIWAKATDSKTREN	216







DE	PUTATIVE POTASSIUM CHANNEL SUBUNIT.
GN	SPCC965.06.
OS	Schizosaccharomyces pombe (Fission yeast).
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;
OC	Schizosaccharomycetes.
NCBI_TaxID=4896;	
RN	[1]
RR	SEQUENCE FROM N.A.
RC	STRAIN=972H-;
RA	Wood V., Rajandream M.A., Barrell B.G., Rieger M.;
RL	Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AL023590; CAA19066.1; -.
KW	Ionic channel.
SQ	SEQUENCE 344 AA; 38430 MW; A57A25E60B05CAE9 CRC64;

  

Query Match	20.9%;	Score 227.5;	DB 3;	Length 344;
Best Local Similarity	32.8%;	Pred. No. 3.5e-10;		
Matches	67;	Conservative 37;	Mismatches 75;	Indels 25; Gaps

  

QY	3	PLEETVYALRDIIVASGKALVYGISGYGPELTSATAAEAFMAEGCPL--LIHQPSYSINRW	60
DB	144	PMEEVVRAFTQLLDGKAFWGTSEWS-AFEIEHAHIATKYNLIAPVADQPQYNLT RD	202
		: :  :	
QY	61	VEPFGDDGENLSAANNGLGVIAFSPLAQGLTLDKYLGIPGGRSASOGKSLSEGMLNV	120
DB	203	HFE-----KDLLPLQQIYGYGATWSPKSGILTKGYNDGIPEGSRSLTFTTSLAGOLQT	257
		: :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
QY	121	----NNIDMYRKUNDIAOERGQSLOAMALAWLREOGEYGDVTVTSAIGASSVEOLDNS	176
DB	258	PEGKTOLDQVRQISKIAEQIATGPSQLALAWTLKN-----PYVSTTIIGASKPEQI---	308
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
QY	177	LDSLNLFESD----AELEAIDEI	196
DB	309	VENVKAVEFDIKLTPEILKKIDEI	332
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	

  

RESULT	12
ID	Q44328 PRELIMINARY; PRT; 354 AA.
AC	Q44328;
DT	01-NOV-1996 (TEMBLrel. 01, Created)
DT	01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DE	01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE	MCCA PROTEIN.
GN	MCCA.
OS	Agrobacterium radiobacter.
OG	Plasmid pTII15955.
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC	Rhizobiaceae; Rhizobium.
OX	NCBI_TaxID=358;
RN	[1]
RR	SEQUENCE FROM N.A.
RC	STRAIN=15955;
RC	MEDLINE=96236046; PubMed=8655509;
RA	Kim K.S., Farrand S.K.;
RT	"Ti plasmid-encoded genes responsible for catabolism of the crown gal-
RT	opine mannopine by Agrobacterium tumefaciens are homologs of the T-
RT	region genes responsible for synthesis of this opine by the plant
RT	tumor.";
RL	J. Bacteriol. 178:3275-3284(1996).
DR	EMBL; AF242881; AAB07785.1; -.
DR	InterPro; IPR001395; -.
DR	Pfam; PF00248; aldo_ket_red; 2.
KW	Plasmid.
SQ	SEQUENCE 354 AA; 38846 MW; 21D4074605E80EB6 CRC64;

  

Query Match	19.5%;	Score 213;	DB 2;	Length 354;
Best Local Similarity	32.5%;	Pred. No. 5.1e-09;		
Matches	69;	Conservative 35;	Mismatches 66;	Indels 42; Gaps

Query Match 19.5%; Score 213; DB 2; Length 354;  
Best Local Similarity 32.5%; Pred. No. 5,1e-09;  
Matches 69; Conservative 35; Mismatches 66; Indels 42; Gaps 9;

```
QY 2 TPLEETMYALRDIVASGKALYGVISY-----GPELTAAAEFMAEEGCPLLIHOP 52
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 134 TPLEETMEALDVLVRQKVRIGCSNFTGWQIMKALGISEKDKRQRFVSQ-----IH-- 186

QY 53 SYSIINRWEEPCDDGNNLQSANNGLVIAFSPLAQGLLTDKYL--DGIPEGSRASOG 110
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 -YTL-----EADAIEYLLPISVDGGLVLIWSPLAGLLSKHRRNOSAPGSRQFAG 239

QY 111 KSL-----SEGMLNVNNDIMVKRLNDIAERGOSLAQMALAWLVRQGEYGDVTSALI 165
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 240 WTEPPYVDEERLWNI-----VDTLLSVADGRGVSAAQVALAWLI-----GRKAVTSIII 288

QY 166 GASSVEQLDNLSDSLNLFESDALEAIDEIS 197
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 289 GGRTEAQFKDNL-AAADLQLSAEERKRLDDVS 319

RESULT 13
Q9HVV1 PRELIMINARY; PRT; 345 AA.
AC Q9HVV1;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PROBABLE OXIDOREDUCTASE.
GN PA4434.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Latbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004857; AAC07822.1; -.
DR InterPro; IPR001395; -.
DR Pfam; PF00248; aldo.ket.red; 1.
DR PRINTS; PR00069; ALDKETRDASE.
DR ProDom; PD002880; -. 1.
SQ SEQUENCE 345 AA; 38644 MW; 6866B69139E62508 CRC64;
```

```
Query Match 19.3%; Score 211; DB 2; Length 345;
Best Local Similarity 33.0%; Pred. No. 7e-09;
Matches 66; Conservative 33; Mismatches 85; Indels 16; Gaps 6;

QY 2 TPLEETMYALRDIVASGKALYGVISY-----GPELTAAAEFMAEEGCPLLIH-QPSYSIINRW 60
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 TPLEETLEVLDEQVRAGKIRHIGLSNETPGMTFLRLAEERGWPRAVSIQNPNLLNS 213

QY 61 VIEPGDDGNNLQSANNGLVIAFSPLAQGLLTDKYL-DGI-PEGSRASQGSLSGMLN 119
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 FE-----VGLAEIATREQGLLAYSPMAGMLSGKYADGARPANARISLYSRFTR-YTN 266

QY 120 VNNIDMVKRLNDIAERGOSLAQMALAWLVRQGEYGDVTSALIGASSVEQLDNLSDS 179
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 PQAEACARYVALAREHGLEPQAALAYVT-----SRPFVTSNIIGATSLQLETNUGS 320

QY 180 LNNLEFSDALEAIDEISH 199
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 V-DLRDLEEVLAGIDAIHRE 339
```

```
RESULT 14
Q9F2Z5 PRELIMINARY; PRT; 330 AA.
AC Q9F2Z5;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE.
GN SCD19.07.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL392149; CAC08308.1; -.
SQ SEQUENCE 330 AA; 36116 MW; 5733CFEF9F7BEEA5 CRC64;

Query Match 19.3%; Score 210.5; DB 2; Length 330;
Best Local Similarity 32.0%; Pred. No. 7.2e-09;
Matches 63; Conservative 36; Mismatches 81; Indels 17; Gaps 6;

QY 2 TPLEETMYALRDIVASGKALYGVISYGPETLTAEEFMAEEG--CPLLIHOPYSIINR 59
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 TPEETIWAQIDVLVQGGKILYAGSNFPGYKTAQANEIARRGGTIGLYSEOCLYNLAER 195

QY 60 WVEEPDGDGNNLQSANNGLVIAFSPLAQGLLTDKYL-DGIPEGSRASQGSLSGMLN 119
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 196 RAE-----MEVIPAARDYGLGVIPWSPHGLGGVKKKATQGRAS-GRA-ADALKD 247

QY 120 VNNIDMVKRLNDIAERGOSLAQMALAWLVRQGEYGDVTSALIGASSVEQLDNLSDS 179
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 248 TKTREQIQAYEDLLDKHGLQPGEAALAWLLTRPG-----VTGPIVGPRTQQLDSALRA 301

QY 180 LNNLEFSDALEAIDEI 196
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 L-ELELSEELSLDEI 317

RESULT 15
Q9L1S0 PRELIMINARY; PRT; 336 AA.
AC Q9L1S0;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE.
GN SCE56.02.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
```



Search completed: September 13, 2001, 17:56:48  
Job time: 298 sec





```

Db 192 -----EEREMPLCYQCGVAVIPNSPLARGLTR-----PWGETTARLVSDVGVKNL 239
QY 114 SEGMLNNDVNRKLNDAQERQSLAQMALAWLVRQEGYADTVTSALIGASSVEQL 173
Db 240 YKES-DENDAQAERTLTGVSEELGATRAQVALAWLLSKFG-----IAAPIIGTSREQL 292
QY 174 DNSLDSLN 181
Db 293 DELLNAVD 300

RESULT 2
ID YDYG_ECOLI STANDARD; PRT; 326 AA.
AC P77256;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE HYPOTHETICAL OXIDOREDUCTASE IN ANSA-RND INTERGENIC REGION.
GN YDYG.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono K., Itoh T., Kanai K., Kasai H.,
RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000272; AAC74841.1; -.
CC EMBL; D90821; BAA15569.1; -.
CC EMBL; D90820; BAA15562.1; -.
CC EcoGene; EG13483; ydJG.
KW Hypothetical protein; Oxidoreductase.
FT ACT_SITE 136 136 HYDROGEN-BOND DONOR (PROBABLE).
SQ SEQUENCE 326 AA; 36328 MW; 1FF5C5252708978 CRC64;

```

```

Query Match 17.7%; Score 193; DB 1; Length 326;
Best Local Similarity 31.0%; Pred. No. 1.6e-08;
Matches 62; Conservative 40; Mismatches 72; Indels 26; Gaps 7;

QY 2 TPLEETWALRDIVASGKALYVGISSYGPETLTAEEAFMAEEGCPILLIHQPSYIINRWV 61
Db 145 TPIAETVAVLNKLSEKGRKIRAGANVDADHIREVLYQY-----GELDIQAKYSILDRAM 199
QY 62 EEPGDDGNLQSAANGLGVAFSPLAQGLTLDKYL-DGIPEGSRASOGKSLSEGMLNV 120

```

```

Db 200 E-----NELLPLCRDNGIWWQVSPLEQGLLTGTITRDVPGGANKVWFQRENMLKV 253
QY 121 NNIDVNRKLNDAQERQSLAQMALAWLVRQEGYADTVTSALIGASSVEQLDNLDSL 180
Db 254 --IDMLEQWQPLCARYQCTIPTLALAWILKQ-----SDLISLGTATAPEQVRENVAAL 305
QY 181 N-NLEFSDAEL-----EAD 194
Db 306 NINLSDADATLREMAEAL 325

RESULT 3
ID YCKC_BACSU STANDARD; PRT; 310 AA.
AC P46905;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE IN NATB-RAPJ INTERGENIC REGION (ORFC).
GN YCKC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kumano M., Tanakoshi A., Yamane K.;
RT "A 32 kb nucleotide sequence from the region of the lincomycin-
RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT identification of the site of the lin-2 mutation.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-268 FROM N.A.
RC STRAIN=BD99;
RA Cheng J., Guffanti A.A., Krulwich T.A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB000617; BAA22238.1; -.
CC EMBL; Z99105; CAB12071.1; -.
CC EMBL; U30873; AAB53024.1; -.
CC HSP; P80276; IAH3.
DR SubtilList; BG11340; yckC.
DR InterPro; IPR001395; -.
DR Pfam; PF00248; aldo_ket_red; 1.
KW Hypothetical protein; Oxidoreductase.
FT ACT_SITE 127 127 HYDROGEN-BOND DONOR (PROBABLE).
SQ SEQUENCE 310 AA; 34800 MW; C06BF4195D25C91C CRC64;

```

```

Query Match 16.7%; Score 182.5; DB 1; Length 310;
Best Local Similarity 29.6%; Pred. No. 1e-07;
Matches 58; Conservative 38; Mismatches 79; Indels 21; Gaps 7;

```

```

QY 2 TPLEETWALRDIVASGKALYVGISSYGPETLTAEEAFMAEEGCPILLIHQPSYIINRWV 61
Db 133 TPLAEVAGTKLKEKDEGKTKAIGASNLQYQ---QLODFNADG--YLEVFQAEYSLIQR-- 185
QY 62 EEPGDDGNLQSAANGLGVAFSPLAQGLTLDKYL-DGIPEGSRASOGKSLSEGMLNV 120
Db 186 ----DAERELLYCEKQGISFIPYPLASGLTGKTFQDTVFDDFR--KDKQFQGETFI 239
QY 121 NNIDVNRKLNDAQERQSLAQMALAWLVRQEGYADTVTSALIGASSVEQLDNLDSL 180

```

Db 240 HNLKVKDKLRAVEEKQADTAHVALWLLTR-----PAIDAIIPGAKRPEQLODNKLT 293  
 QY 181 NNLFEADAELEFAIDEI 196  
 Db 294 -NIELTEDEVNFISDI 308

## RESULT 4

TAS\_ECOLI ID TAS\_ECOLI STANDARD; PRT; 346 AA.  
 AC Q46933; 1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE TAS PROTEIN.  
 GN TAS.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B/R / WU3610;  
 RX MEDLINE=98240727; PubMed=9560382;  
 RA Timms A.R., Bridges B.A.;  
 RT "Reversion of the tyrosine ochre strain Escherichia coli WU3610 under  
 starvation conditions depends on a new gene tas.";  
 RL Genetics 148:1627-1635 (1998).  
 CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U29581; AAB40481.1; -;  
 DR EMBL; AE000367; AAC75873.1; -;  
 DR EMBL; Y14609; CAA74961.1; -;  
 DR HSP; P14550; 2ALR.  
 DR EcGene; BG13093; tas.  
 DR InterPro; IPR001395; -;  
 DR Pfam; PF00248; aldo\_ket\_red; 1.  
 KW Oxidoreductase.  
 FT ACT\_SITE 132 132 HYDROGEN-BOND DONOR (BY SIMILARITY).  
 SQ SEQUENCE 346 AA; 38499 MW; 111692D06CA07CD7 CRC64;

Query Match 16.5%; Score 180.5; DB 1; Length 346;  
 Best Local Similarity 30.1%; Pred. No. 1.7e-07;  
 Matches 63; Conservative 42; Mismatches 63; Indels 41; Gaps 10;

QY 4 LEETMYALRDIVASGKALYGVIS----SYGPDLTAEEAFMAEEGCPLLIH-QPSYSIINR 59  
 Db 157 LLDTDLALAEYQKIRYIGVSNFAGVMRYLHLAD---KHLPRIVTQNPYSILNR 213  
 QY 60 WVEPGDDGNNLQSAANGLGVTAFAFSLAQLTLDKYLGI-PEGSRAS-----QG 110  
 Db 214 SFE-----VGLAEVSQYEGVELLAYSCGLGLTGKLNKAPGARGNTLFSRFTYSG 267

QY 111 KSLSEGMLNNVNDIMVRKLNIDIAQERGQSLAQMALAWLVRQEGYGADTVTSALIGASSV 170  
 Db 268 EOTCKA-----VAAYVDIARRHGLDPAQMALAFVRQG-----PFVASTLLGATTM 312  
 QY 171 EQDNLSDLSLNNLEFSD---AELEAIDEI 196  
 Db 313 DQKLTNIESL-HLELSEVDLAEIEAVHQV 340

## RESULT 5

YAKC\_SCHPO ID YAKC\_SCHPO STANDARD; PRT; 340 AA.  
 AC Q09923;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE HYPOTHETICAL 37.7 KDA PROTEIN CIF7.12 IN CHROMOSOME I.  
 GN SPAC1F7.12 OR SPAC21E11.01.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Murphy L., Niblett D., Harris D., Barrell B.G., Rajandream M.A.,  
 RA Walsh S.V.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; Z67998; CAA91959.1; -;  
 DR EMBL; Z67999; CAA91961.1; -;  
 DR HSP; P14550; 2ALR.  
 DR InterPro; IPR001395; -;  
 DR Pfam; PF00248; aldo\_ket\_red; 1.  
 KW Hypothetical protein; Oxidoreductase.  
 FT ACT\_SITE 126 126 HYDROGEN-BOND DONOR (PROBABLE).  
 SQ SEQUENCE 340 AA; 37710 MW; A4A9686A70968F06 CRC64;

Query Match 16.3%; Score 178; DB 1; Length 340;  
 Best Local Similarity 26.8%; Pred. No. 2.7e-07;  
 Matches 55; Conservative 42; Mismatches 88; Indels 20; Gaps 8;

QY 1 DTPLEETMYALRDIVASGKALYGVISSYGPDLTAEEAFMAEEGCPLLIHQPSYSIINRW 60  
 Db 131 ETPIEKIMGALKKCKVEAGKIRYIGLSECSANTIRRAAVY-----PVSAVQVEYSPFSL 185  
 QY 61 VEEPGDDGNNLQSAANGLGVTAFAFSLAQLTLDKYL--DGIPEG-SRASQSKSLSEGM 117  
 Db 186 IERP---EIGVMKACRENNITIVCYAPLGRGLTGAKYKSPDFEGDFRKPAPYQENF 242  
 QY 118 LVNNVNDIMVRKLNIDIAQERGQSLAQMALAWLVRQEGYGADTVTSALIGASSVBQLDNSL 177  
 Db 243 --YKMLEVTKIEKIATANNITPGQLSLAWLLAQ---GDDIL--PIPGTKRVKYLENF 294  
 QY 178 DSLNNLEFSDAELEAIDEISHDAGI 202  
 Db 295 GAL-KVKLSDATVKEIREACDRAEV 318

## RESULT 6

YM98\_MYCTU ID YM98\_MYCTU STANDARD; PRT; 323 AA.

AC Q50668;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HYPOTHETICAL 34.9 KDA PROTEIN RV2298.  
 GN RV2298 OR WTCY339.12C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37BV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Horsby T., Jorgels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 CC -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z77163; CAB00973.1; -  
 DR HSP: P03023; ILCD.  
 DR TubercuList; RV2298; -  
 KW Hypothetical protein; Oxidoreductase.  
 FT ACT\_SITE 117 117 HYDROGEN-BOND DONOR (BY SIMILARITY).  
 SQ SEQUENCE 323 AA; 34986 MW; FF071A48B745B8E8 CRC64;  
 -----  
 Query Match 14.8%; Score 161.5; DB 1; Length 323;  
 Best Local Similarity 32.8%; Pred. No. 5.4e-06;  
 Matches 61; Conservative 29; Mismatches 71; Indels 25; Gaps 6;  
 QY 8 MYALRDIVAGKALYVGISSYGPDLTAFAAEFMAEGCPILLIHQPSYSIINRWVEPGDD 67  
 DB 129 MPMRLDLSGGDIGAGVSNYS---LAWRKADAALGRPVSNQVHFSLAH-----PDA 179  
 QY 68 GENLQSAANNGILGVIAFSPLAGGLTIDKY-LDGIPEGSRSQGSLSSEGLNVDW 126  
 DB 180 LEDLVFAELNRIVAYSPLAGGLGGYGLNREGGVRLNPLEGTENLRRIE--PLL 237  
 QY 127 RKLNDIAQREGSLQAMALAWLVREGYGVADVTTSALIGASSVEQLDLSLNLEFS 186  
 DB 238 ATLRAIVDVAQPAQVALWLSLPG-----VVAIFGASSVEGLE-----FNVA 284  
 QY 187 DALEA 192  
 DB 285 DIELSA 290  
 RESULT 7  
 ID AR72\_HUMAN STANDARD; PRT; 330 AA.  
 AC Q43488; O75749;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE AFLATOXIN B1 ALDEHYDE REDUCTASE 1 (EC 1.1.1.1) (AFB1-AR 1)  
 DE (ALDOKETOREDUCTASE 7).

GN AKR7A2 OR AFAR OR AKR7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=98244807; PubMed=9576847;  
 RA Ireland L.S., Harrison D.J., Neal G.E., Hayes J.D.;  
 RT "Molecular cloning, expression and catalytic activity of a human AKR7  
 RT member of the aldo-keto reductase superfamily: evidence that the  
 RT major 2-carboxybenzaldehyde reductase from human liver is a homologue  
 RT of rat aflatoxin B1-aldehyde reductase."  
 RL Biochem. J. 332:21-34(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99040634; PubMed=9823300;  
 RA Praml C., Savelyeva L., Perri P., Schwab M.;  
 RT "Cloning of the human aflatoxin B1-aldehyde reductase gene at 1p35-  
 RT 1p36.1 in a region frequently altered in human tumor cells."  
 RL Cancer Res. 58:5014-5018(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC Hall R.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: CAN METABOLIZE THE AFLATOXIN B1 (AFB1) PROTEIN-BINDING  
 CC DIHYDRODIOL BY FORMING NONBINDING AFB1 DIALCOHOL. COULD BE  
 CC INVOLVED IN PROTECTION OF LIVER AGAINST THE TOXIC AND CARCINOGENIC  
 CC EFFECTS OF AFLATOXIN B1. ACTS AS A 2-CARBOXYBENZALDEHYDE  
 CC REDUCTASE.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF026947; AAC52104.1; -  
 DR EMBL: Y16675; CAA76347.1; -  
 DR EMBL: AL035413; CAB72321.1; -  
 DR MIM: 603418; -  
 KW Oxidoreductase.  
 FT ACT\_SITE 112 112 HYDROGEN-BOND DONOR (PROBABLE).  
 FT CONFLICT 113 113 A -> T (IN REF. 1).  
 SQ SEQUENCE 330 AA; 36618 MW; 3BBFB7ED0CAF4D54 CRC64;  
 -----  
 Query Match 14.6%; Score 159.5; DB 1; Length 330;  
 Best Local Similarity 27.8%; Pred. No. 8e-06;  
 Matches 59; Conservative 33; Mismatches 95; Indels 25; Gaps 7;  
 QY 2 TPLEETMYALRDIVAGKALYVGISSYGPDLTAFAAEFMAEGCPIL-LIHQPSYSIINRW 60  
 DB 118 TPVEETLHACQRLHQEGKVELGLSNYSASWEVAEICTLCKSNWILPTVYOGYNATTQ 177  
 QY 61 VEEPGDDGENLQSAANNGILGVIAFSPLAGGLTIDKYLDGIPEGSRSQ---GKLSSEG 117  
 DB 178 VE-----TELPFCLRHFRFYAVNPLAGGLLTGKYKYEDKDGKVPVGRFFGNSWAET 231  
 QY 118 LN-----VNIDMVRLNDIAQ-ERGOSLAQMALAWLVRE---OGEYGVADVTTSALIG 166  
 DB 232 RNRFWKEHFEALVEKALQAYGASVTSALRWYHHHSQLOCAHG----DAVLG 287  
 QY 167 ASSVEQLDLSLNLEFSDAELEAIDEISH 198  
 DB 288 MSSLEQLQNLAATBEGPLEPAVVDVAFNQA 319

Query Match	14.3%;	Score 156.5;	DB 1;	Length 327;
Best Local Similarity	26.3%;	Pred. No. 1.4e-05;		

Matches 57; Conservative 37; Mismatches 94; Indels 29; Gaps 8;

QY 2 TPLEETMYALRDIVASKALYGVIGSYGPELTAAAEFMAEEGCP-LIHOPYSIINRW 60  
Db 115 TPIETLOACHHHVHQBCKFVGLGSLNVSVEVAEICTLCKKNWIMPTVYQGMYNALTRQ 174  
QY 61 VEPGDDGNNLQSAANNGLGVIAFSPLAQGLLTDKYLDGIPGSRASQ---GKSLSEGM 117  
Db 175 VE-----TELFPCRLRHFLGRFYAFNPLAGGLLTGRYKQDKGNPESRFFGNPFSQLY 228  
QY 118 L-----NVNNDIMYRK-LNDIAQERGQSLAQMALAWLRE---OGEYGADTVTSALIG 166  
Db 229 MDRYKEEHFNGALVEKALKTTGYTPASMSIAARVYVHHSQLKGTQG----DAVILG 284  
QY 167 ASSVEQLDNLSDLNLEFSDAELEIDE-----ISHD 199  
Db 285 MSSLEQLQNALVEGCPLEPAVVADFQAWNLVAHE 321

RESULT 10  
STCV\_EMENI STANDARD; PRT; 387 AA.  
AC Q00727;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS DEHYDROGENASE STCV  
GN (EC 1.1.1.-).  
DE STCV.  
OS Emericella nidulans (Aspergillus nidulans).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; Emericella.  
OX NCBI\_TaxID=5072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FGSC 26;  
RX MEDLINE=96202293; PubMed=8643646;  
RA Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,  
RT "Twenty-five coregulated transcripts define a sterigmatocystin gene  
cluster in Aspergillus nidulans."  
RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).  
CC -!- FUNCTION: MAY BE INVOLVED IN THE DEHYDRATION OF 5'-  
HYDROXYAVERANTIN TO FORM AVERUFIN.  
CC -!- PATHWAY: STERIGMATOCYSTIN BIOSYNTHESIS.  
CC -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U34740; AAC49206.1; -  
KW Oxidoreductase.  
FT ACT\_SITE 148 HYDROGEN-BOND DONOR (POTENTIAL).  
SQ SEQUENCE 387 AA; 43548 MW; 1621588273B85E8C CRC64;

Query Match 14.0%; Score 152.5; DB 1; Length 387;  
Best Local Similarity 27.0%; Pred. No. 3.6e-05;  
Matches 57; Conservative 33; Mismatches 80; Indels 41; Gaps 8;

QY 2 TPLEETMYALRDIVASKALYGVIGSYGPELTAAAEFMAEEGCP-LLIHOPYSIINRW 60  
Db 154 TSVEEYMQSLHLYVACKVNLNGISDAPAWVAKCMYARFHLGRFCVQI-----GRW 207  
QY 61 VEPGDDGNNLQSAANNGLGVIAFSPLAQGLLTDKYLDGIPGSRASQKSL-----SEC 116  
Db 208 ACSYRDFEREILLPMQCEGLALAPWALGRG-----QYKSAEEPFQEG 250

QY 117 MLNVNND-----MVRKLNIDIAQERGQSLAQMALAWLREOGEYGADTVTSALIGASSVE 171  
Db 251 TRNNGPQEEKRLMKAGKLTVEGKERGVAAAIALAYLLHK-----SPYFVPVIGCRTVE 304  
QY 172 QLDNSLSDLNLEFSDAEL-EAIDEISHDAG 201  
Db 305 QLEANITSL-GVELSDEEIEIEDTIPFDVG 334

RESULT 11  
AL15\_TOBAC STANDARD; PRT; 307 AA.  
ID AL15\_TOBAC  
AC P40691;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE AUXIN-INDUCED PROTEIN PCNT115.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
OC Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. WHITE BURLEY;  
RL van der Zaai E.J.;  
RL Submitted (AUG-1990) to the EMBL/GenBank/DBJ databases.  
CC -!- INDUCTION: BY AUXIN.  
CC -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X56267; CAA39708.1; -  
DR PIR; S16390; S16390;  
DR InterPro; IPR001395; -  
DR Pfam; PF00248; aldo\_ket\_red; 1.  
KW Oxidoreductase.  
FT ACT\_SITE 136 HYDROGEN-BOND DONOR (PROBABLE).  
SQ SEQUENCE 307 AA; 33857 MW; C4417DA852613360 CRC64;

Query Match 13.7%; Score 150; DB 1; Length 307;  
Best Local Similarity 25.8%; Pred. No. 4.2e-05;  
Matches 49; Conservative 36; Mismatches 59; Indels 46; Gaps 7;

QY 3 PLEETMYALRDIVASKALYGVIGSYGPELTAAAEFMAEEGCPLLIHOPYSIINRWYE 62  
Db 143 PLEITVGLKLLVEEGLKYIGLS-----EASASTIR-RAHAVHPITAVQLEWSLRSRDE 197  
QY 63 EPGDDGNNLQSAANNGLGVIAFSPLAQGLLTDKYLDGIPGSRASQKSLSEGM LN--- 119  
Db 198 -----EEIIPCTRELIGIGIVAYSP LGRGFL-----SSGPKLLDEMSNEDY 237  
QY 120 -----VNNIDMYRKLNIDIAQERGQSLAQMALAWLREOGEYGADTVTSALIGA 167  
Db 238 RYLPFRQFQENLNKNLYERICEMAVRKGTSPQLAWLWHHQ-----GND---VCDIPGT 291  
QY 168 SSVEQLDNL 177  
Db 292 TKIENLNQNM 301

RESULT 12  
YOKF\_BACSU STANDARD; PRT; 306 AA.  
ID YOKF\_BACSU  
AC P54569;  
DT 01-OCT-1996 (Rel. 34, Created)



```
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL OXIDOREDUCTASE IN ANSR-BMRU INTERGENIC REGION.
GN YQKF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=168 / JH642;
RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RA Sato T., Takeuchi M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D84432; BAA12638.1; -
DR EMBL; Z99116; CAB14294.1; -
DR Subtilist; BG11761; Yqkf.
DR InterPro; IPR001395; -
DR Pfam; PF00248; aldo_ket_red; 1.
DR Hypothetical protein; Oxidoreductase.
FT ACT_SITE 119 119 HYDROGEN-BOND DONOR (PROBABLE).
SQ SEQUENCE 306 AA; 34717 MW; CD79B13B9568000F CRC64;

Query Match 13.4%; Score 146; DB 1; Length 306;
Best Local Similarity 25.58; Pred. No. 8.9e-05;
Matches 50; Conservative 44; Mismatches 74; Indels 28; Gaps 8;

QY 4 LEETMYALRDIVASGKALYGVISYGPPELTAEAEFMAEAGCPPLIHQPSYSIINRWEE 63
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
127 IDEITAEFEELKQEGVIRYGISSIRPNV---IKYVKKSIVSIMQ--FSLFDRRPEE 181

QY 64 PGDDGENLQSAANGLGVIAFSPLAGLLTDKYLDGIPGSRASQKSLG-EGMLNVNN 122
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
182 -----WLPLEBHQISVVGARGVAKGLTEKPLDQASESMKONGYLSFEELTNAR- 233

QY 123 IDMYRKLNDIAQERGQSLAQMALAWLREOGEYCADVTSTALIGASSVEQLDNLSD 182
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
234 ----KAMEEVAPD--LSMTEKSLQYLLAQ-----PAVASVITGASKIEQLRENIQAANA 281

QY 183 LEFSDAELEAIDEISH 198
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
282 RRLTEEEIKALQ--SH 295

RESULT 13
IDLS_BACSU STANDARD; PRT; 310 AA.
AC P46336;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE IOLS PROTEIN (VEGETATIVE PROTEIN 147) (VEG147).
GN IOLS OR SS92ER.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=168 / BGSC1A1;
RA Yoshida K.-I., Seki S., Fujimura M., Miwa Y., Fujita Y.;

RT "Cloning and sequencing of a 36-kb region of the Bacillus subtilis
RT genome between the gnt and iol operons.";
RL DNA Res. 2:61-69(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=168 / BGSC1A1;
RA Fujita Y., Shibayama T., Ishio I., Aoyama D., Yoshida K.-I.;
RT "Organization and transcription of the myo-inositol operon, iol, of
RT Bacillus subtilis.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-20.
RC STRAIN=IS58;
RX MEDLINE=97443988; PubMed=9298659;
RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
RA Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
RT regulation map for Bacillus subtilis.";
RL Electrophoresis 18:1451-1463(1997).
CC -1- PATHWAY: MYO-INOSITOL CATABOLISM.
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB005554; BAA21607.1; -
DR EMBL; Z99124; CAB16014.1; -
DR HSSP; P23457; ILWI.
DR Subtilist; BG11363; iols.
DR Pfam; PF00248; aldo_ket_red; 1.
KW Oxidoreductase.
FT ACT_SITE 125 125 HYDROGEN-BOND DONOR (PROBABLE).
SQ SEQUENCE 310 AA; 35168 MW; A870F226F8684867 CRC64;

Query Match 13.4%; Score 146; DB 1; Length 310;
Best Local Similarity 25.18; Pred. No. 9e-05;
Matches 49; Conservative 37; Mismatches 83; Indels 26; Gaps 7;

QY 2 TPLEETMYALRDIVASGKALYGVISYGPPELTAEAEFMAEAGCPPLIHQPSYSIINRW 61
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
131 TPKDEAVNALNEMKACKIRSIGVSNFSLQKEA---NKDGL-VDVLOGEYNLLNREA 185

QY 62 EEPGDDGENLQSAANGLGVIAFSPLAGLLTDKYLDG--IPEGSRASQKSLSEGLN 119
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
186 E-----KTFPPYTKHNISFIPYPLVSLGGLAGKYTEDTTFPEGLRNEQEHF-KGERF 238

QY 120 VNNIDMYRKLNDIAQERGQSLAQMALAWLREOGEYCADVTSTALIGASSVEQLDNLSD 179
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
239 KENIRKYNKLAIAERKHNDIPHIVLAWYL-----ARPEIDILIPGAKRADQL----- 286

QY 180 LNNLEFSDAELEAID 194
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
287 IDNKTADVLSQED 301

RESULT 14
GS69_BACSU STANDARD; PRT; 331 AA.
AC P80874; Q07583;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE GENERAL STRESS PROTEIN 69 (GSP69).
GN YHDN.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
```

```
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-25.
RC STRAIN=1558;
RX MEDLINE=97443988; PubMed=9298659;
RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
RA Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
RL regulation map for Bacillus subtilis.";
RL Electrophoresis 18:1451-1463(1997).
CC -!- INDUCTION: BY HEAT SHOCK, SALT STRESS, OXIDATIVE STRESS, GLUCOSE
CC LIMITATION AND OXYGEN LIMITATION.
CC -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Y14082; CAAT74498.1; -
DR EMBL; 299109; CAB12792.1; -
DR HSSP; P23457; ILWI.
DR Subtilisin; BG13020; yhdN.
DR Pfam; PF00248; aldo_ket_red; 1.
KW Oxidoreductase; Heat shock.
FT ACT_SITE 125 125 HYDROGEN-BOND DONOR (PROBABLE).
FT CONFLICT 25 25 G -> K (IN REF. 2).
FT SEQUENCE 331 AA; 37312 MW; 82BC24D46E4994D0 CRC64;
CC -----
Query Match 13.3%; Score 145.5; DB 1; Length 331;
Best Local Similarity 26.1%; Pred. No. 0.00011;
Matches 57; Conservative 37; Mismatches 91; Indels 33; Gaps 8;
QY 3 PLEETMYALRDIVASGKALYVGISSYGPPELTAAAEFAAEAGCPLLIHQPSYSIINRWVE 62
DB 132 PIETAEVVKELVDAGKIRAGVSNFSIE---QMDTFRAV--APLHTIQPPVNLFRERE 186
QY 63 EPDGDGENLQSAANGLGVIAFSPLAQGLLTDKYL-----DGIPEGSRASOGKSLSE 115
DB 187 -----ESVLPYAKDKNITLLYGLRCGLLTCKMTTEYTFEGDDLRNHPKQKPRFE 240
QY 116 GMLNVNMDVMKLNDAERGOSLAQMALAWVLRQEGEYGDVTSALIGASSVQDLN 175
DB 241 YLSAVNQLDKLAK-----TRYGKSVIHLAVRWILDQP---GADI---ALWGARKPGOLEA 289
QY 176 SLD-----SLNNLEFSDAELEADIEISHDAGINIWAAT 209
DB 290 LSEITGWTLNSEDDQKDINTILENTISDPVGPPEFMAPPT 327
RESULT 15
NORA ASPFL
ID NORA ASPFL STANDARD; PRT; 388 AA.
AC Q00049;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NORSOLORINIC ACID REDUCTASE (EC 1.1.1.-).
GN NORA OR ADH-2.
OS Aspergillus flavus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5059;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 3357;
RX MEDLINE=96156784; PubMed=8593042;
RA Cary J.W., Wright M., Bhatnagar D., Lee R., Chu F.;
RT "Molecular characterization of an Aspergillus parasiticus
RL dehydrogenase gene, nora, located on the aflatoxin biosynthesis gene
RN cluster.";
RL Appl. Environ. Microbiol. 62:360-366(1996).
CC -!- PATHWAY: AFLATOXIN BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U32377; AAC49167.1; -
KW Oxidoreductase.
FT ACT_SITE 148 148 HYDROGEN-BOND DONOR (POTENTIAL).
FT SEQUENCE 388 AA; 43751 MW; 758B8187187BF5B0 CRC64;
SQ
Query Match 12.8%; Score 139.5; DB 1; Length 388;
Best Local Similarity 25.1%; Pred. No. 0.00041;
Matches 49; Conservative 39; Mismatches 84; Indels 23; Gaps 7;
QY 2 PLEETMYALRDIVASGKALYVGISSYGPPELTAAAEFAAEAGCPLLIHQPSYSIINRW 60
DB 154 TSVEEYMRSLNHLVANGKVLGLVSDTPAWLVVYKNAFARANGLTPFSVYQGHWSCAFR- 212
QY 61 VEEPGDGENLQSAANGLGVIAFSPLAQGLLTDKYLDGIPGSRASOGKSLSEGMLNV 120
DB 213 -----DFERDILPMCESEGWGLAPWGLVGRG-----QFRSAEEFSR--EGRKM--GPQDE 258
QY 121 NNIDVMVKLNDAERGOSLAQMALAWVLRQEGEYGDVTSALIGASSVQDLNLSL 180
DB 259 KHRRLEKLDQMAQKQKTRKATSIQAQYVMHKA-----PYVFPVIGRKVEHLKENIEAL 312
QY 181 NNLEFSDAELEAIDE 195
DB 313 -GLVLSEEEIREIDD 326
Search completed: September 13, 2001, 17:57:09
Job time: 294 sec
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:34:55 ; Search time 25.14 Seconds  
(without alignments)  
654.483 Million cell updates/sec

Title: us-09-105-117i-2  
Perfect score: 1091  
Sequence: 1 DTPLEETMYALRDIVASGKA.....SHDAGINIWAKATDSKTREN 216

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	529.5	48.5	346	2 G65086	hypothetical prote
2	526.5	48.3	346	2 E85959	probable reductase
3	449	41.2	319	2 T35337	probable ion chann
4	366	33.5	287	2 B86873	oxidoreductase yug
5	283.5	26.0	315	2 A75289	probable potassium
6	267	24.5	319	2 H72391	K+ channel, beta s
7	257	23.6	336	2 T44988	oxidoreductase [im
8	238	21.8	336	2 C84256	probable oxidoredu
9	237.5	21.8	387	2 F82815	voltage-gated pota
10	227.5	20.9	344	2 T41659	probable potassium
11	225	20.6	324	2 E85538	probable NAD(P)H-d
12	225	20.6	348	2 C64771	probable oxidoredu
13	213	19.5	354	2 T44934	moCA protein limpo
14	211	19.3	345	2 B83093	probable oxidoredu
15	209	19.2	330	2 T07394	probable potassium
16	209	19.2	367	2 A53131	alpha-dendrotoxin-
17	208	19.1	367	2 S66502	potassium channel
18	208	19.1	367	2 S45312	RCKbeta2 protein -
19	206.5	18.9	316	2 F83170	probable oxidoredu
20	205	18.8	401	2 S66503	potassium channel
21	205	18.8	408	2 I59393	potassium channel
22	205	18.8	419	2 I55463	K+ channel beta-su
23	197.5	18.1	326	2 D85787	hypothetical prote
24	197	18.1	352	2 C82294	oxidoreductase Tas
25	196	18.0	324	2 D84315	aryl-alcohol dehyd
26	195	17.9	328	2 T52133	potassium channel
27	194.5	17.8	316	2 F69978	sugar-phosphate de
28	193	17.7	326	2 C64937	hypothetical prote
29	193	17.7	327	2 E75296	aldo/keto reductas

ALIGNMENTS

RESULT 1

G65086

hypothetical protein b3001 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence,revision 17-Sep-1997 #text\_change 18-Feb-2000

C:Accession: G65086

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: G65086

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-346 <BLAT>

A:Cross-references: GB:AE000382; GB:U00096; NID:g2367182; PIDN:AC76037.1; PID:g17893

A:Experimental source: strain K-12, substrain MG1655

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match	48.5%	Score	529.5	DB 2	Length	346			
Best Local Similarity	51.4%	Pred. No.	3.1e-34						
Matches	110	Conservative	38	Mismatches	51	Indels	15	Gaps	3
QY	1	DTPLEETMYALRDIVASGKALYGVISYGPPELTAEAEFMAEGCPLLIHQPSYSIINRW	60						
Db	143	NTPMEETASALAHAVQSGKALYGVISYSPERTQKMWELLRWKIPLLIHQPSYNLLNRW	202						
QY	61	VEEPGDGGENLQSAANNGLGVIAFSPLAQGLLTDKYLDGIPESGR----	ASQGSLSSEG	116					
Db	203	VDKSG-----LLDTLQNNVGVCIAFTPLAQGLLTGKYLINGIPQDSRMHREKNGKVLTPK	257						
QY	117	MLVNNITDMYRKNDIAQERGQSLAQMALVLRQEGYGDVTSALIGASSVVEQLDNS	176						
Db	258	MLTEANLNLRLNEMAAQQRQSQMAQSLWLLK-----DRVTSVLIGASRAEQLEEN	311						
QY	177	LDSLNNEFFSDAELEAIDEISHDAGINIAKATD	210						
Db	312	VQALNNLTFTSKELAQIDQIHADGELNLWQASSD	345						
RESULT	2								
E85959									
probable reductase 24354 [imported]									
C:Species: Escherichia coli									
C>Date: 16-Feb-2001 #sequence,revision 16-Feb-2001 #text_change 31-Mar-2001									
C:Accession: E85959									
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May									
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda									
Nature 409, 529-533, 2001									
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.									
A:Reference number: A85480; MUID:21074935; PMID:11206551									

A:Accession: E85959

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-346 &lt;STO&gt;

A:Cross-references: GB:AE005174; NID:g12517567; PIDN:AAG58137.1; GSPDB:GNO0145; UWGP:243

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z4354

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match

Best Local Similarity 48.3%; Score 526.5; DB 2; Length 346;

Matches 109; Conservative 39; Mismatches 51; Indels 15; Gaps 3;

Db 143 NTPEETASALAHAVOSGKALYGVISYSPERTOKKVELLHEWKIPLLIHOPSYNLLNRW 202

QY 61 VEEPGDDGENLQSAANNGLVIAFSPLAQGLTDRKYLDPGEGSR----ASQCKSLSEG 116

Db 203 VDKSG-----LLDTLQNNVGCCIAFPLAQLGLTGKYLNGIPEDSRMRHREGNKVRLTPK 257

QY 117 MLVNNIDVYRKNDIAQERGQSLAQMALAWLVREOGEYADVTTSALIGASSVEQLDNS 176

Db 258 MLTEANLNSRLNEMAAQOQSGMAQMSLWLLKD-----ERTSVLVGASRAEQLEEN 311

QY 177 LDSLNLEFSDAELEAIDEISHDAGINIAKATD 210

Db 312 VQALNNLTFTSELAQIDQIADGENLWQASSD 345

RESULT 3

T35337

probable ion channel subunit - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 18-Feb-2000

C:Accession: T35337

R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

A:Reference number: Z21575

A:Accession: T35337

A:Molecule type: DNA

A:Residues: 1-319 &lt;OLI&gt;

A:Cross-references: EMBL:AL049863; PIDN:CAB42946.1; GSPDB:GNO0070; SCODEB:SC5H1.21c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCODEB:SC5H1.21c

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match

Best Local Similarity 41.2%; Score 449; DB 2; Length 319;

Matches 105; Conservative 25; Mismatches 54; Indels 12; Gaps 4;

QY 2 TPLEETMYALRDIVASGKALYGVISYSPELTAFAAEFMAEEGCPPLIHOPSYSIINRW 61

Db 132 TPLEETMGALHTAVQOGKALYGVSVNSPEQTREAAAILAELGTPLLIHQPRYSMLDRP 191

QY 62 EEPGDDGENLQSAANNGLVIAFSPLAQGLTDRKYLDPGEGSR-ASQCKSLSEGMLNV 120

Db 192 ETGG-----LLDALDELEVGSYSPLEQGLTGRYLNGIPEDSRASDSPFLNSDAVTE 246

QY 121 NNIDVYRKNDIAQERGQSLAQMALAWLVREOGEYADVTTSALIGASSVEQLDNSL 180

Db 247 ELVGKRLGRLNEIATAGRSQSLAQLAWVL-R-QGR-----VTSALVGASSPRQLEDSVAI 300

QY 181 NNLEFSDAELEAIDEI 196

Db 301 GNLDFADELARIDKI 316

RESULT 4

B86873

oxidoreductase yugB [Imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 31-Mar-2001

C:Accession: B86873

R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh

Genome Res. in press, 2001

A:Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625

A:Accession: B86873

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-287 &lt;STO&gt;

A:Cross-references: GB:AE005176; NID:g12725031; PIDN:AAK06084.1; GSPDB:GNO0146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: yugB

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match

Best Local Similarity 33.5%; Score 366; DB 2; Length 287;

Matches 83; Conservative 22; Mismatches 37; Indels 10; Gaps 5;

QY 1 DTPEETMYALRDIVASGKALYGVISYSPELTAFAAEFMAEE-GCPPLIHOPSYSIINR 59

Db 143 NTPIETMGALKSALDSGKALYVGLSNYSAKET-EAAVLAALKFKLLIHQPRYSMLDR 201

QY 60 WVEPGDDGENLQSAANNGLVIAFSPLAQGLTDRKYLDPGEGSRASOG--KSLSEGM 117

Db 202 WIE---DDLQETLSEA---GIGIAFKPLYQGLLTKYLHGIPEDSRMRDPHYATLHDDS 255

QY 118 LNVNNDIVYRKNDIAQERGQSLAQMALAWVL 149

Db 256 LTKKRLQEQVALNDLAQSRGQSLAQMALAWVL 287

RESULT 5

A75289

probable potassium channel, beta subunit - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C:Accession: A75289

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: A75289

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-315 &lt;WHI&gt;

A:Cross-references: GB:AE002063; GB:AE000513; NID:g6460121; PIDN:AAF11861.1; PID:g646

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR2317

A:Map position: 1

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match

Best Local Similarity 26.0%; Score 283.5; DB 2; Length 315;

Matches 68; Conservative 40; Mismatches 76; Indels 13; Gaps 4;

QY 1 DTPEETMYALRDIVASGKALYGVISYSPELTAFAAEFMAEEGCPPL-LIHOPSYSIINR 59

Db 127 EYPMEEIVMAFDQVTRDGRKALYWGTSWMPAARIAQAQAVEFAKANGLHAPYTEQPEYSWVR 186

QY 60 WVEPGDDGENLQSAANNGLVIAFSPLAQGLTDRKYLDPGEGSRASQCKSLSEGM 119

Db 187 DRVE-----QEILPYTEGAGIGLVVWSPLAMGLLTCKYDEGREGARLTKENWAGSYLT 241

```

A;Note: expressed during exponential growth
C;Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match                23.6%; Score 257; DB 2; Length 336;
Best Local Similarity      32.1%; Pred. No. 7.5e-13;
Matches 67; Conservative 40; Mismatches 84; Indels 18; Gaps

QY      1  DTPLEETALRDIVASSKALYGVIGSSYPPEL--TAAEAEFMAERGC-PLLIHQPSYSII 57
      :||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      141  EPTREMKLTNGLVEDKHYLGASTLRPNAMKVARANEIARAEGWEPFTVAQPRYNLV 200

QY      58  NRWVEPQDGENLQSAANGNGLVAFSPLAGLLTDXYL--DGIPGSRASOGKSLSE 115
      :||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      201  DREIE--GD---YLEMTRSYGIAVCPWSPGLGCGFLTGYKTYREDGLGTGESRAAESRRFEE 254

QY      116  GMLNVNNTDMYRKLNDTAQERGOSLAQMALAWLYRQGEYGADTVTSALIGASSVEOLDN 175
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      255  SYLTENEFDVHDELDVAGEVDATPAQTALAMLHRDG-----VTAPIVGARTVEQLTE 308

QY      176  SLDSLNNEFSDAELEAIDETSHDAGINI 204
      :||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      309  NLEAA-TIDLDEQVDRLTGAKEDPYVGL 336

RESULT      8
C84256
probable oxidoreductase [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2000
C;Accession: C84256

```

```

Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.;
A>Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: AB4160; MUID:20504483
A:Accession: C84256
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <STO>
A:Cross-references: GB:AE004437; NID:g10580553; PIDN:AAG19415.1; GSPDB:GNC000000000
C:Genetics:
C:Gene: yajO2
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match          21.8%; Score 238; DB 2; Length 336;
Best Local Similarity 33.5%; Pred. No. 2.3e-11;
Matches 67; Conservative 34; Mismatches 81; Indels 18; Gaps

QY    1   DTPLEETWALRDVAVSGKALYVGISYGPEL--TAAEAEPMAEEGC-PLLIHOPSYSII 57
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    141 DTARELMRTINDLVADGLVHYHGCASTLPQNAWRIRANELARAGEWEFFSVLPQRYNLV 200

QY    58 NRWVEPGDGDGENLQSAANNNGLVIAFSLQAQLLTDKYL--DGIPGSRASOGKSLSE 115
      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    201 DREIE--GD----YLTFAREQGLAVCPWSPLAQGFLLTKGYSRDDCLGTCKSKASDSRRWD 256

QY    116 GMLNVNNIDMWVKLNDTAQERGOSLAOMALAWVLVRQGEYGAADVTSALICASSVEQLDN 175
      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    255 TYLTANDTAHDHVLDADAVERASPQVALWLLHRDG-----ITAPIVGARTPDOLSE 308

QY    176 SLDSLNNLFSDAELEAIDE 195
      ::| | : : : : : :
Db    309 NIDAA-TIDLTDQDMARLTD 327

RESULT      9
FB2815
C:Species: Xylella fastidiosa [imported] - Xylella fastidiosa

```

F82813  
 voltage-gated potassium channel beta subunit XF0367 [imported] - *Xylella fastidiosa* (C:Species: *Xylella fastidiosa*)

```

Query Match      20.9%, Score 227.5; DB 2; Length 344;
Best Local Similarity 32.8%; Pred. No. 1.6e-10;
Matches 67; Conservative 37; Mismatches 75; Indels 25

QY      3 PLEETVVALRDIVASGKALYGVISSYGPETLTAFAAFAEAGCPL--LIHQPSYSYISII
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      144 PMEEVVRAFTLQDGKAFYNGTSEWS-AFEIEHAHIAIKYLNLIAPVADQPYNYI
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      61 VEFGDGDENLQSANNGLGVIAERSPLAQGLLTDRKYLDGIPGSRASQGSKSLSEGG
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      203 HFE-----KDLLPQQIYGYGATVMSPLKSGILTKGYNDGIPGSRSLTFTTSLAGG
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      121 ---NNIDMRKLNIDIAERGOSLAQMAWLVLREOGEYGADVTLSALICASSVVEQ
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      258 PEGKTLDQVQVQISKIAEQIATPQSOLALAWTLKN-----PVSTTILGASKEPQ
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      177 LDSLNLEFSD----AELEAIDEI 196
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      309 VENVKAVEFDIKLTPEILKKIDEI 332
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
E85538
Probable NAD(P)-H-dependent xylose reductase yajO [imported] - Esche
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31
C:Accession: E85538
R:Ferna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.H.;
  Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Po
  Nature 409, 529-533, 2001
A:title: Genome sequence of enterohemorrhagic Escherichia coli O157
A:Reference number: E855480; MUID:21074935; PMID:11206551

```

Query Match 20.6%; Score 225; DB 2; Length 324;  
Best Local Similarity 31.9%; Pred. No. 2.3e-10;  
Matches 60; Conservative 39; Mismatches 63; Indels 26;

Qy	60	WVEEFGDGDGENLLGSAANGLGVIAFSPLAQGLLTKYLDGIPEGRASO-----G
Db	192	-----EEERMLPLCYOEQVAVIPWSPLARGRLTR-----PWGETTARLVSDVGV
Qy	114	SEGLMNVNNIDMVRKLNIDTAQERGOSGAQMALAWLREGEYGADVTVSALIGASSVSS
Db	240	YKES--DENDAQIAERLTGVSEELGATRAQVALAWLLSKPG-----IAAPIICTSRRE
Qy	174	DNSLDSLNL 181
Db	293	DELLNAVD 300
 RESULT 12		
C64771		
probable oxidoreductase (EC 1.-.-.) yajO - Escherichia coli		
C:Species: Escherichia coli		
C:date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 18-		
C:Accession: C64771		
R:Blattnet, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burlat		
A.; Rose, D.J.; Mau.B.; Shao, Y		



QY 111 KSL-----SEGMLNVNNIDMVRKLNIDIAQERQSLAQMALAWVLRQGEYGADTVTSALI 165





[illegible]

US-08-576-626A-31  
 ; Sequence 31, Application US/08576626A  
 ; Patent No. 5998194  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Summers, R.G.  
 ; APPLICANT: Katz, L.  
 ; APPLICANT: Donadio, S.  
 ; APPLICANT: Staver, M.J.  
 ; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR  
 ; TITLE OF INVENTION: BIOSYNTHESIS GENES  
 ; NUMBER OF SEQUENCES: 60  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Abbott Laboratories  
 ; STREET: 100 Abbott Park Road  
 ; CITY: Abbott Park  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60064-3500  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/576,626A  
 ; FILING DATE: 21-DEC-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Dianne Casuto  
 ; REGISTRATION NUMBER: P-40,943  
 ; REFERENCE/DOCKET NUMBER: 5857.US.O1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (847) 938-3137  
 ; TELEFAX: (847) 938-2623  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 31:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1114 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: No. 5998194e  
 ; US-08-576-626A-31

Query Match 16.0%; Score 175; DB 2; Length 1114;  
 Best Local Similarity 29.9%; Pred. No. 2.3e-10;  
 Matches 59; Conservative 38; Mismatches 80; Indels 20; Gaps 7;  
 QY 3 PLEETMYALRDIVASGKALYVGISYGPGLTAAAEFMA-EEGCPLLIHOPSYIINRW 61  
 Db 919 PWDEVQAMPSLVASGKSYVGSNFGWIAAQAENARRHSLGVMYSHOCLYNLVRHA 978  
 QY 62 EEPGDDGNNLQSAANGLGVAFSPLAQGLLTDKYLDGIPESG--RASOGKSLSEGMLN 119  
 Db 979 E-----LEVLPAQAAYGLGVFAMSPHLGGLLSGA-LEKLAAGTAVKSAQGRA---QVLL 1028  
 QY 120 VNNIDMYRKLNDAIQAERGOSIAQMALAWLREOGEYADTVTSALIGASSVEQLDNLSDS 179  
 Db 1029 PSURPAEAYEKFCRNGEDPAEGLAWLVRPG-----IAGAVIGPRTPEQLDSALKA 1082  
 QY 180 LNNLEFSDAELEAIDEI 196  
 Db 1083 -SMTLDEQALSELDEI 1098

RESULT 5  
 US-08-907-674-1  
 ; Sequence 1, Application US/08907674  
 ; Patent No. 5919685

; GENERAL INFORMATION:  
 ; APPLICANT: Bandman, Olga  
 ; APPLICANT: Shah, Purvi  
 ; APPLICANT: Guegler, Karl J.  
 ; APPLICANT: Corley, Neil C.  
 ; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE  
 ; NUMBER OF SEQUENCES: 3  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/907,674  
 ; FILING DATE: Herewith  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J.  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0362 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-855-0555  
 ; TELEFAX: 415-845-4166  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 331 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: BRAINOT14  
 ; CLONE: 1596452  
 ; US-08-907-674-1  
 Query Match 14.9%; Score 162.5; DB 2; Length 331;  
 Best Local Similarity 27.8%; Pred. No. 8e-10;  
 Matches 59; Conservative 34; Mismatches 94; Indels 25; Gaps 7;  
 QY 2 TPLEETMYALRDIVASGKALYVGISYGPGLTAAAEFMAEEGCPPL-LIHOPSYIINRW 60  
 Db 119 TPVEETLHACORLHQEGCKFVELGSLNSYASWEVAEICTLCKSNWILPTVYOGMYNAITRQ 178  
 QY 61 VEPPGDDGNNLQSAANGLGVAFSPLAQGLLTDKYLDGIPESGRASQ---GKSLSEG 117  
 Db 179 VE-----TELPCLRFHGLRFYAYNPLAGLLTGKYEKDKGKQPVGRFFGNTWAEY 232  
 QY 118 LN-----VNNIDMYRKLNDAIQAERGOSIAQMALAWLRE---QGEYGADTVTSALIG 166  
 Db 233 RNRYWKHEHFEGLALVEKALQAAYGASAPSVTSAAALRWYHHSOLOQAHG-----DAVILG 288  
 QY 167 ASSVEQLDNLSDLSNNLEFSDAELEAIDEISH 198  
 Db 289 MSSLEQLEQNLAATEEGPLEPVAVVDFAPNQAWH 320

RESULT 6  
 US-09-215-087-1  
 ; Sequence 1, Application US/09215087  
 ; Patent No. 5981244  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bandman, Olga

```

; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/215,087
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/907,674
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0362 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT14
; CLONE: 1596452
;
US-09-215-087-1

```

```

Query Match 14.9%; Score 162.5; DB 2; Length 331;
Best Local Similarity 27.8%; Pred. No. 8e-10;
Matches 59; Conservative 34; Mismatches 94; Indels 25; Gaps 7;

QY 2 TPLEETWYALRDIVASGKALYGVISYGPETAEAEFNAEGCPL-LIHQPSYSIINRW 60
Db 119 TPVEETLHACQRLHQRKGKVELGSLNSYASWEVAEICTLCKSNGLWILPTVYQGMYNATIRQ 178
QY 61 VEEPGDGGNLLQSAANGLGVIAFSPLAGGLTDKYLGIPEGRASQ---GKSLSEGM 117
Db 179 VE-----TELPCLRHFGLRYAYNPLAGGLTGKYYEDKDGKPGVGRFGNTWAEMY 232
QY 118 LN-----VNNIDMVRKLNIDIAQ-ERGOSLAQMALAWLRLRE---QGEYGADTVTSALIG 166
Db 233 RNRVKEHHEFGIALVEKALQAAYGASAPSVTSALRWMYHHSQLOQAHG----DAVILG 288
QY 167 ASSVEOLDNSLDLNNLEFSDAELEAIDEISH 198
Db 289 MSSLEQLEONLAATEEGPLEPAVVDFAFNQAWH 320

```

```

RESULT 7
US-09-391-959-1
; Sequence 1, Application US/09391959
; Patent No. 6071704
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.

```

```

; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/391,959
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/907,674
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0362 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT14
; CLONE: 1596452
;
US-09-391-959-1

```

```

Query Match 14.9%; Score 162.5; DB 3; Length 331;
Best Local Similarity 27.8%; Pred. No. 8e-10;
Matches 59; Conservative 34; Mismatches 94; Indels 25; Gaps 7;

QY 2 TPLEETWYALRDIVASGKALYGVISYGPETAEAEFNAEGCPL-LIHQPSYSIINRW 60
Db 119 TPVEETLHACQRLHQRKGKVELGSLNSYASWEVAEICTLCKSNGLWILPTVYQGMYNATIRQ 178
QY 61 VEEPGDGGNLLQSAANGLGVIAFSPLAGGLTDKYLGIPEGRASQ---GKSLSEGM 117
Db 179 VE-----TELPCLRHFGLRYAYNPLAGGLTGKYYEDKDGKPGVGRFGNTWAEMY 232
QY 118 LN-----VNNIDMVRKLNIDIAQ-ERGOSLAQMALAWLRLRE---QGEYGADTVTSALIG 166
Db 233 RNRVKEHHEFGIALVEKALQAAYGASAPSVTSALRWMYHHSQLOQAHG----DAVILG 288
QY 167 ASSVEOLDNSLDLNNLEFSDAELEAIDEISH 198
Db 289 MSSLEQLEONLAATEEGPLEPAVVDFAFNQAWH 320

```

```

RESULT 8
US-08-907-674-3
; Sequence 3, Application US/08907674
; Patent No. 5919685
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE

```

```

, , NUMBER OF SEQUENCES: 3
, , CORRESPONDENCE ADDRESS:
, , ADDRESSEE: Incyte Pharmaceuticals, Inc.
, , STREET: 3174 Porter Drive
, , CITY: Palo Alto
, , STATE: CA
, , COUNTRY: USA
, , ZIP: 94304
, , COMPUTER READABLE FORM:
, , MEDIUM TYPE: Diskette
, , COMPUTER: IBM Compatible
, , OPERATING SYSTEM: DOS
, , SOFTWARE: FastSeq for Windows Version 2.0
, , CURRENT APPLICATION DATA:
, , APPLICATION NUMBER: US/08/907,674
, , FILING DATE: Herewith
, , CLASSIFICATION: 514
, , PRIOR APPLICATION DATA:
, , APPLICATION NUMBER:
, , FILING DATE:
, , ATTORNEY/AGENT INFORMATION:
, , NAME: Billings, Lucy J.
, , REGISTRATION NUMBER: 36,749
, , REFERENCE/DOCKET NUMBER: PF-0362 US
, , TELECOMMUNICATION INFORMATION:
, , TELEPHONE: 415-855-0555
, , TELEFAX: 415-845-4166
, , TELEX:
, , INFORMATION FOR SEQ ID NO: 3:
, , SEQUENCE CHARACTERISTICS:
, , LENGTH: 327 amino acids
, , TYPE: amino acid
, , STRANDEDNESS: single
, , TOPOLOGY: linear
, , IMMEDIATE SOURCE:
, , LIBRARY: GenBank
, , CLONE: 433611
, , US-08-907-674-3

```

```
Query Match      14.3%; Score 156.5; DB 2; Length 327;  
Best Local Similarity 26.3%; Pred. No. 3.6e-09;  
Matches          57; Conservative   37; Mismatches    94; Indels     29; Gaps       8;
```

	QY	2	TPIEETVVALRDIVASKALYVGITSSYGPELTAAEAEEFMAEGCPL-LIHOPYSIIINRW	60
Dbb	:::::	:	: : :: : :: :	:: :: :: : : :: :
	115	TPIEETLQAHHVHQEKGFVELGLSNVSVEVAEICTLCKNGHIMPTVYOGMNATRQ	174	
QY	61	VIEPFGDDGNLLSAAANGLGVIAPSLAQGLTKDYLDGIGEPSRASQ----	GKSLSBQM	117
Dbb	:: ::	: : :: : :: :	: : ::	:
QY	175	VE-----TELFPCRHFRLRFYAANLAGLLTGRYKYQDKDKNBPESFFGNPFSQLY	228	
Dbb	L-	-----NYNNIDMWPK-LNDIAQRQSILAOMAWLVRE----OGEYGADTVTSALIG	166	
QY	229	MDRYWKEEHENGATLEBAUKTTVPTRAPSISAARWRYMHSOLKGTHQG-----DAVLIG	284	
Dbb				
QY	167	ASSVEQLDNDSLNNEFEFSDAELEAIDE-----ISHD	199	
Dbb	285	MSSLDEONLALEEGLPEPAVVVDAGPDOWNLVAHE	321	

RESULT 9  
US-09-215-087-3  
; Sequence 3, Application US/09215087  
; Patent No. 5981244  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/215,087  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/907,674  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0362 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 327 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 433611  
US-09-215-087-3

Query Match	14.3%	Score 156.5;	DB 2;	Length 327;
Best Local Similarity	26.3%;	Pred. No. 3.6e-09;		
Matches	57;	Conservative	37;	Mismatches 94;
			Indels	29;
			Gaps	8;
QY	2	TPLEETVYALRDIVASKALVVGTSYSGPELTAAEAERMAEEGCL-LIHOPSYSLINRW	60	
DB	115	TPLEETVYALRDIVASKALVVGTSYSGPELTAAEAERMAEEGCL-LIHOPSYSLINRW	174	
QY	61	VEEPDGDGNLLOSAANNLGVIAFSPLAQGLLTDKYLDGIPESRASQ-----GKSLSEGM	117	
DB	175	VE-----TELPCLRHFGLRFYAFNPLAGLLTGRYKQDKGKNPESRFGNPFPSQLY	228	
QY	118	L-----NVNNIDMVRK-LNDIAQERGOSLAQMAWLALRE-----OGEYGADTVTSALIG	166	
DB	229	MDRYWKEEHENGIALVEKALTKTYTGPAPTSMISAARVMYHHSQLKGTQ-----DAVILG	284	
QY	167	ASSVEQLDNLSDLSLNNLEFSDAELEAIDE-----ISHD	199	
DB	285	MSSLEOLEONLALVEEGPLEPVPVADGPDQAWNVAHE	321	

RESULT 10  
US-09-391-959-3  
; Sequence 3, Application US/09391959  
; Patent No. 6071704  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive

CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/391,959  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/907,674  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0362 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 327 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 433611  
US-09-391-959-3

Query Match 14.3%; Score 156.5; DB 3; Length 327;  
Best Local Similarity 26.3%; Pred. No. 3.6e-09;  
Matches 57; Conservative 37; Mismatches 94; Indels 29; Gaps 8;  
QY 2 TPLEETMYALRDIVASGKALYGVISYGPDLTAFAAEFMAEEGCPL-LIHQPSYSILNRW 60  
Db 115 TPLEETLQACHHHVQGGKFEVLSNYSWEVAEICTLCKKNGWIMPTVYQGMYNITRQ 174  
QY 61 VEPFGDDGNNLSAANNGVIAFSPLAGGLTDKYLDCIPEGRASQ---GKSLSEG 117  
Db 175 VE-----TELPCLRHFGRLFFAFNPLAGGLTGRYKQYQDKGKNPESRFFGNPESQLY 228  
QY 118 L-----NVNNDIMVRK-LNDIAQERGOSLAOMALAWLRE---GQEGYADTVTSALIG 166  
Db 229 MDYRWKEEHFNGLAVEKALKTTTGPTAPSMISAARVMYHHSQKGTGQ----DAVILG 284  
QY 167 ASSVEQLDNLSDNLNLEFSDAELEAIDE---ISHD 199  
Db 285 MSSLEQLQNLALVEEGPLEPAVVDADFQANLVAHE 321

RESULT 11  
US-09-166-412-4  
Sequence 4, Application US/09166412  
Patent No. 6133504  
GENERAL INFORMATION:  
APPLICANT: Wayne D. Loescher,  
APPLICANT: John D. Everard  
APPLICANT: Rebecca Grumet  
TITLE OF INVENTION: DNA Encoding Mannose  
TITLE OF INVENTION: 6-Phosphate Reductase  
TITLE OF INVENTION: and Recombinants  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ian C. McLeod  
STREET: 2190 Commons Parkway

CITY: Okemos  
STATE: Michigan  
COUNTRY: USA  
ZIP: 48864  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 5.25 inch,  
MEDIUM TYPE: 360 kb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/166,412  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/731,320  
FILING DATE: October 15, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ian C. McLeod  
REGISTRATION NUMBER: 20,931  
REFERENCE/DOCKET NUMBER: MSU 4.1-275  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (517) 347-4100  
TELEFAX: (517) 347-4103  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310  
TYPE: Amino Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE:  
DESCRIPTION:  
HYPOTHETICAL: No  
ANTI-SENSE: No  
ORIGINAL SOURCE:  
ORGANISM: apple  
IMMEDIATE SOURCE:  
LIBRARY: N/A  
US-09-166-412-4

Query Match 9.9%; Score 107.5; DB 4; Length 310;  
Best Local Similarity 20.8%; Pred. No. 0.00091;  
Matches 44; Conservative 47; Mismatches 72; Indels 49; Gaps 7;  
QY 4 LEETMYALRDIVASGKALYGVISYGPDLTAFAAEFMAEEGCPL-LIHQPSYSILNRWEE 63  
Db 139 LOOTWEGMEKTVSLGLVRSIGLSNYELFUTRDC-----LAYSIIKPAVSQFETH 187  
QY 64 PGDDGNNLSAANNGVIAFSPLAGGLTDKYLDCIPEGRASQKSLSEGMLNVNII 123  
Db 188 PYFQDRLVKFCMKHGVLPHTAHTPLG-GAANKDMFG-----SVSPL 228  
QY 124 DMVRKLNDAQERGOSLAOMALAWLREGEYGAADTVTSALIGASVEQLDNLSDNLN 183  
Db 229 D-DPVLNDVAKKYKGSVAQICLRW-----GIQKRTAVIPKSKKIQRLENLEVL-EF 278  
QY 184 EFSDALEAIDEI-----SHDAGINIWA 206  
Db 279 QLSDEDMQLIYSIDRKRYRTSLPSKWTGLDYYA 310

RESULT 12  
US-09-166-412-2  
Sequence 2, Application US/09166412  
Patent No. 6133504  
GENERAL INFORMATION:  
APPLICANT: Wayne D. Loescher,  
APPLICANT: John D. Everard  
APPLICANT: Rebecca Grumet  
TITLE OF INVENTION: DNA Encoding Mannose  
TITLE OF INVENTION: 6-Phosphate Reductase



RESULT 13  
US-08-336-198C-3  
; Sequence 3, Application US/08336198C  
; Patent No. 5866382

[illegible]

```

RESULT 14
US-09-222-817-12
; Sequence 12, Application US/09222817
; Patent No. 6037154
; GENERAL INFORMATION:
; APPLICANT: MIKIKO SUGA, Masatazu S
; TITLE OF INVENTION: METHOD OF PRODUCING
; FILE REFERENCE: OP813
; CURRENT APPLICATION NUMBER: US/09/222-817-12
; CURRENT FILING DATE: 1998-12-30

```

```

Query Match      8.8%; Score 95.5; DB 3; Length 530;
Best Local Similarity 23.6%; Pred. NO. 0.047;
Matches 59; Conservative 30; Mismatches 106; Indels 55; Gaps

QY      10 ALRDIVASGKALYVCISYGPGLTAAAEFAEBAECPLLIIHQPSYIINRWVEPG--DD 67
      || | || | | | | | | | | | | | | | | | | | | | | | | |
Db      243 ALADAIESGHRGAGFDVYSTEPCTDSPLFK----LPQWVTPHLGASTEAAQDRAGTDV 298

QY      68 GENILQSANN-----GLGIYASPLAQ--GLITDKYLDGIPGSRASQCKS 112
      : : : : : | | | | | | | | | | | | | | | | | | | | | |
Db      299 ADSVLKALAGEFVADAVNVSGRGVEKVAVMMDLARKLGLLGLVDAAPYVSIEVEARGE 358

```

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:51:44 ; Search time 34.57 Seconds  
(without alignments)  
508.561 Million cell updates/sec

Title: US-09-105-117I-3

Perfect score: 1460

Sequence: 1 MNPIDTLISIDEGSFG.....RSLRLTDAVDAIEGLRP 290

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0601.\*

- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1460	100.0	290	AAW37714	C. glutamicum Lys
2	1460	100.0	290	AAW37714	Corynebacterium gl
3	118.5	8.1	278	AAW39317	ORF116 protein inv
4	105	7.2	311	AAW36129	Snpr activator pro
5	105	7.2	311	AAW30494	Amino acid sequenc
6	102.5	7.0	299	AAW74592	Neisseria meningit
7	101.5	7.0	299	AAW74593	Neisseria meningit
8	100.5	6.9	344	AAW29277	Protein encoded by
9	99.5	6.8	1346	AAW18640	Amino acid sequenc
10	99.5	6.8	1346	AAW77195	S. venezuelae macr
11	99.5	6.8	1346	AAW67204	Narbonolide syntha

12	99.5	6.8	12199	21	AAW77180	S. venezuelae pik
13	98	6.7	303	21	AAW74591	Neisseria gonorrhoe
14	95.5	6.5	5069	19	AAW52846	A. mediterranei ri
15	93.5	6.4	298	13	AAW25588	Truncated irgB. v
16	92.5	6.3	928	21	AAW77292	Streptomyces antib
17	92.5	6.3	928	21	AAW78844	Ksq-Atg loading di
18	92.5	6.3	4150	21	AAW92707	S. antibioticus 8,
19	92	6.3	401	20	AAW33264	E. coli biosl prot
20	92	6.3	401	20	AAW33268	Plasmid pHS1 biosl
21	92	6.3	401	20	AAW33270	DE19731274 Seq ID
22	92	6.3	401	20	AAW92938	DE19731274 Seq ID
23	92	6.3	401	20	AAW92936	E. coli biotin ORF
24	92	6.3	401	20	AAW92934	Cephalosporin C #2
25	89.5	6.1	383	15	AAW49827	S. venezuelae pik
26	89.5	6.1	1346	21	AAW77203	Bud promoter trans
27	89	6.1	290	17	AAW02378	S. clavuligerus OR
28	87.5	6.0	432	16	AAW77865	Sorangium cellulos
29	87.5	6.0	3798	21	AAW58577	Regulatory protein
30	87	6.0	303	21	AAW80094	Bovine p57 protein
31	86.5	5.9	461	17	AAW98342	Cellulose synthase
32	86.5	5.9	1319	14	AAW45002	Neisseria meningit
33	86	5.9	391	21	AAW95652	Human ORF2255
34	84.5	5.8	3266	21	AAW42491	Amino acid sequenc
35	84.5	5.8	3739	21	AAW18638	S. venezuelae macr
36	84.5	5.8	3739	21	AAW77193	S. venezuelae pik
37	84.5	5.8	3739	21	AAW77201	Narbonolide syntha
38	84.5	5.8	3739	21	AAW67202	Protein encoded by
39	83.5	5.7	2188	22	AAW66467	S. antibioticus 8,
40	83.5	5.7	3816	21	AAW92708	Cephalosporin anti
41	83	5.7	419	12	AAW10691	Phosphoenolpyruvat
42	83	5.7	856	17	AAW96949	A. mediterranei ri
43	83	5.7	1688	19	AAW52848	SpnB a polyketide
44	82.5	5.7	2152	20	AAW39298	Tylactone synthase
45	82.5	5.7	4472	18	AAW22601	

## ALIGNMENTS

RESULT 1  
AAW37714  
ID AAW37714 standard; Protein; 290 AA.  
XX  
AC AAW37714;  
XX  
DT 12-MAR-1998 (first entry)  
XX  
DE C. glutamicum Lys G protein (lysine export regulator).  
XX  
KW LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;  
KW Microbial production; amino acid; animal feed additive.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN DE19548222-AL.  
XX  
PD 26-JUN-1997.  
XX  
PF 22-DEC-1995; 95DE-1048222.  
XX  
PR 22-DEC-1995; 95DE-1048222.  
XX  
(KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.  
XX  
PA Eggeling L, Sahm H, Vrljic M;  
XX  
PI WPI; 1997-333867/31.  
XX  
DR N-PSDB; AAT96816.  
XX  
PT Increasing microbial production of amino acids, especially lysine -  
PT by improving export carrier activity or corresponding gene  
PT expression, also new export and regulatory genes from  
PT Corynebacterium



microlides which are useful for the control of arachnids, nematodes and insects. Biosynthesis of spinosyns occurs via stepwise condensation and modification of carboxylic acid precursors generating a linear polypeptide which is modified further. The DNA sequence contains a central region of approximately 55kb which has homology to the DNA encoding the polypeptide synthases (PKS) of known macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with stop codons at the end of acyl carrier protein (ACP) domains. Together the PKS polypeptides (AAV39297-V39301), form a complex consisting of an initiator module, spnA, and several extender modules spnB-spnE. The products of the genes present in the region upstream of the PKS genes have been assigned names spnF-spnS AAV39302-V39315 and are responsible for different modifications in spinosyn biosynthesis. There are also two ORFs ORFL15 and ORFL16 present immediately upstream of spnS, producing polypeptides AAV39316-V39317, and two ORFs ORP1 and ORP2 present downstream of the PKS region producing polypeptides AAV39318-V39319. It is suggested that the ORFL16 protein is involved in transcriptional control during spinosyn biosynthesis. The genes are useful to improve yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis, or interruption of steps in spinosyn biosynthesis. The modified spinosyns may be a new insect control agent or serve as substrates for further chemical modification and the creation of new semi-synthetic spinosyns. The genes are also useful to isolate similar sequences from *S. spinosa* or other species by hybridization.

```

Query Match      8.1%; Score 118.5; DB 20; Length 278;
Best Local Similarity 24.5%; Pred. No. 0.00015;
Matches 71; Conservative 47; Mismatches 139; Indels 33; Gaps 11;

QY 1 MNPIQLDILLSIIDEGSFEGASIALSISPSAVSORKALEHHVGRVLVSRP-QPAKATEA 59
Db 1 metrelryfvavaeelhfrgaarqlgiaaqplsrliatqleqlgvllqrtsrkvsltea 60

QY 60 GEVLVQARKMV--LLOAETHKAQLSGRLAEIPLTIANADSLTSTWFFPVFNEVASWGAT 117
Db 61 gamlltegrailgalaaeertrgraa-tspsvlvlaakagagsgellakildayaeeapav 119

QY 118 LTLRLDEAHTLSLLR--RGDVLGAVTRRANPNVACGEVVELGTMRLHIAITPSSLURDAYM 175
Db 120 avdillcesgqktilhgradv-allhqfptaealdieilnteqqvailpts----- 171

QY 176 DGKLDWAAMPVLREGPKDVLQDRDLDCRGVDCPGRRRVSIVPSAEG-----FGEAIR 227
Db 172 ---hplasephvmadvsslpdlpl-arwpgdg-----vypdpggvevrnqtqlfgmla 222

QY 228 RGLGWGLLPETQAPMLKA-GEVILLDEIPIDTPMYQWRRLSRSLARL 276
Db 223 lgrttvympeessvnlleglaavpvlpadgvttvialwop-hsrslralaql 271

```

Db	172	---	hplasephvrmadvsslpdlpl-arwppdg-----vypdpgvevrnqtqlfgmla	222
QY	228	RGLCWGLLPETOQAAPMLKA-GEVILLDEIPIDTPMYQWRWLESGSLARL	276	
		:   :	:   :	
Db	223	lgrttvmpessrvnlleglaavpvdapdvttviawpp-hrsrslagl	271	
		:   :	:   :	
	RESULT	4		
	AAW36129			
	ID	AAW36129 standard; Protein; 311 AA.		
	XX			
	XX	AAW36129;		
	XX			
	21-MAY-1998	(first entry)		
	DT			
	XX			
	XX	SnpR activator protein.		
	DE			
	XX			
	XX	Daucomycin C-14 hydroxylase; doxA gene; doxorubicin; daunomycin;		
	KW	13-dihydrocarinomycin; carminomycin; anthracycline;		
	KW	anticancer; cytostatic; cancer; therapy; plasmid pANT195; SnpR.		
	XX			
	OS	Streptomyces sp.		
	XX			
	Key	Location/Qualifiers		
	FH	18..39		
	Domain	/note= "alpha-helix-beta-turn-alpha helix		
	FT	DNA-binding domain"		
	FT			

AA	key	Location/Qualifiers
FF	FF	18..39
FT	Domain	/note= "alpha-helix-beta-turn-alpha helix
FT		DNA-binding domain"
FT		

```

XX PN WO9744439-A2.
XX PD 27-NOV-1997.
XX PF 22-MAY-1997; 97WO-US08690.
XX PR 24-MAY-1996; 96US-0653650.
XX PA (OHIS ) UNIV OHIO STATE. RES FOUND.
XX PI DeSanti CL, Dickens ML, Strohl WA;
XX DR WPI: 1998-018495/02.
XX DR N-PSDB; AAV01451.
XX PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
XX PT -also hydroxylation and oxidation of other anthracycline(s) with
XX PT the same enzyme
XX PS Disclosure; Fig 6; 59pp; English.
XX CC This protein is the encoded product of the Streptomyces snpr
XX CC activator gene. The snpr activator gene is incorporated in
XX CC novel plasmid pANT195 (see AAV01451) that also includes the
XX CC Snpr-activated snpA promoter and a modified doxA gene (see
XX CC AAV01447) of Streptomyces sp. strain C5. The doxA gene codes for
XX CC daunomycin C-14 hydroxylase (see AAW36128), an enzyme capable of
XX CC converting daunomycin to the anticancer agent doxorubicin. Host
XX CC cells, especially Streptomyces host cells, transformed with pANT195
XX CC can be used in methods for the production of doxorubicin from
XX CC daunomycin or for the hydroxylation and oxidation of other
XX CC anthracyclines.
XX SQ Sequence 311 AA;

Query Match 7.2%; Score 105; DB 19; Length 311;
Best Local Similarity 21.2%; Pred. No. 0.0052;
Matches 66; Conservative 50; Mismatches 129; Indels 66; Gaps 13;

QY 6 LDTLLSTIDGSEFEGASLALSISPSAVSORVKALEHHVGRVLSRTQP-AKATEAGEVLV 64
DB 8 lralcaiaadgslhrraarglvtqpslstqrriehelgalfvrtartgcrptplgrlvi 67
QY 65 QAARKMWL-----LQATTKAQLSGRLAEIPLTIAINAD-SLSTWFFPPVFNEVASWGATLT 119
DB 68 srarplvaelcslvseara---aavadaslrvgstasralagw----lrlrhwdqptlh 120
QY 120 LRLEDEAHTLSLLRRGDVLGAVTREPANPVAGCEVVELGTMRHLAIATPSLRDAYMWVDGKL 179
DB 121 mdvsana-llrmvadhghdvafrvehe-----gsir-----vpeglrvrvlvqrep 165
QY 180 DWAAAMPVLRFGPKDVLQDRDLGRVDPVGR-----RRVSIVPSAEGFGGAIRR 228
DB 166 qfvcpl-----adhpaaeatasyaptwttrmidptvdgwnavrr 207
QY 229 -----GLGWGLL--PETQAAPMLKAGEVILLDE--IPIDTPMYQWRWLRSLRLTDA 279
DB 208 vlraegldsrilghdyhsaayvatgevtvqvqptsraetaavrllhgdplgrvlllaa 267
QY 280 VDAATEGLRP 290
DB 268 rtdtelegvyp 278

RESULT 5
AAB30494
ID AAB30494 standard; Protein; 311 AA.
XX
AC AAB30494;
XX
XX 06-MAR-2001 (first entry)

```

```

XX DE Amino acid sequence of Snpr encoded by plasmid pANT1201.
XX KW Streptomyces sp. strain C5; SnpA; S. venezuelae; alpha-amylase;
XX KW endostatin; cancer; tumour growth; angiogenesis.
XX OS Streptomyces sp.
XX PN WO200060945-A1.
XX PD 19-OCT-2000.
XX PF 12-APR-2000; 2000WO-US09747.
XX PR 13-APR-1999; 99US-0129084.
XX PA (MERI ) MERCK & CO INC.
XX PI Desanti CL, Strohl WR;
XX DR WPI: 2000-686970/67.
XX DR N-PSDB; AAC62024.
XX PT Preparation of soluble recombinant endostatin involves transforming
XX PT Streptomyces host with expression vector comprising nucleotide
XX PT sequence encoding endostatin operably linked to linker and leader
XX PT peptide
XX PS Example 1; Fig 9A-B; 57pp; English.
XX CC The present sequence is encoded by a fragment of pANT1201. This plasmid
XX CC was used as a source of snpr, snpA promoter, and multiple cloning site.
XX CC The specification describes a method for the production of soluble,
XX CC recombinant human endostatin in Streptomyces. Leader sequences of
XX CC Streptomyces sp. strain C5 SnpA and S. venezuelae alpha-amylase proteins
XX CC are linked to the N-terminal of endostatin. This ensures that endostatin
XX CC protein is produced as a secreted, soluble protein which needs no
XX CC refolding, is stable in the fermentation broth and is produced in large
XX CC quantities. The method is used for preparing soluble recombinant human,
XX CC murine or primate endostatin, which is useful in the treatment of cancer,
XX CC inhibition of tumour growth, inhibition of angiogenesis, isolation of
XX CC receptors for endostatin and for identification of anti-angiogenic
XX CC compounds in assays. The endostatin protein is produced as a secreted,
XX CC soluble protein which needs no refolding, is stable in the fermentation
XX CC broth and is produced in large quantities. Streptomyces are amenable
XX CC for cultivation in large fermentations allowing for large quantities of
XX CC soluble endostatin to be produced.
XX SQ Sequence 311 AA;

Query Match 7.2%; Score 105; DB 21; Length 311;
Best Local Similarity 21.2%; Pred. No. 0.0052;
Matches 66; Conservative 50; Mismatches 129; Indels 66; Gaps 13;

QY 6 LDTLLSTIDGSEFEGASLALSISPSAVSORVKALEHHVGRVLSRTQP-AKATEAGEVLV 64
DB 8 lralcaiaadgslhrraarglvtqpslstqrriehelgalfvrtartgcrptplgrlvi 67
QY 65 QAARKMWL-----LQATTKAQLSGRLAEIPLTIAINAD-SLSTWFFPPVFNEVASWGATLT 119
DB 68 srarplvaelcslvseara---aavadaslrvgstasralagw----lrlrhwdqptlh 120
QY 120 LRLEDEAHTLSLLRRGDVLGAVTREPANPVAGCEVVELGTMRHLAIATPSLRDAYMWVDGKL 179
DB 121 mdvsana-llrmvadhghdvafrvehe-----gsir-----vpeglrvrvlvqrep 165
QY 180 DWAAAMPVLRFGPKDVLQDRDLGRVDPVGR-----RRVSIVPSAEGFGGAIRR 228
DB 166 qfvcpl-----adhpaaeatasyaptwttrmidptvdgwnavrr 207
QY 229 -----GLGWGLL--PETQAAPMLKAGEVILLDE--IPIDTPMYQWRWLRSLRLTDA 279

```



Query Match	7.0%;	Score 102.5;	DB 21;	Length 299;
Best Local Similarity	26.3%;	Pred. NO. 0.0092;		
Matches 49;	Conservative	31;	Mismatches 67;	Indels 39;
Matches 49;	Conservative	31;	Mismatches 67;	Gaps 7;

CC Neisserial bacteria (e.g.

CC Neisserial bacteria (e.g.

CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.

XX SQ Sequence 299 AA;

Query Match 7.0%; Score 101.5; DB 21; Length 299;  
 Best Local Similarity 26.9%; Pred. No. 0.012; Mismatches 30; Indels 39; Gaps 8;  
 Matches 50; Conservative 30; Mismatches 67; Indels 39; Gaps 8;  
 QY 2 NPIQLTLLSIIDEGSFEGASLALISPSAVSQRVKALEHHVGRVLSR-TOPAKATEAG 60  
 Db 4 nseeltvfvvsgsfraaeqlamansavsrivkrleeklgvnlrrttqlslsteeg 63  
 QY 61 EYLVQARWVLLQAEKTKAQLSGRLAEIPLTTAINADSLSTWPPV-----FNE- 109  
 Db 64 agyfrraqrlqemaaetemia-vheipqgv-lrvdsa---mpmvihlilaplaakfner 118  
 QY 110 -----VASWGGATLTURLLEDAHTLSLLRRGDVLTREANPVGCEVVELGTMRH 161  
 Db 119 yphirlslvssegynliierkvdia-----lragel-----ddsglrarhlfdsrf 164  
 QY 162 LAIATP 167  
 Db 165 rviasp 170

RESULT 8  
 AAY29277  
 ID AAY29277 standard; Protein; 344 AA.  
 XX AC AAY29277;  
 XX DT 25-OCT-1999 (first entry)  
 XX DE Protein encoded by ORF2.  
 XX KW Human pathogen; virulence polypeptide; virulence factor;  
 XX KW pathogenic infection; Pseudomonas aeruginosa infection.  
 XX OS Pseudomonas aeruginosa.  
 XX PN W0927129-A1.  
 XX PD 03-JUN-1999.  
 XX PF 25-NOV-1998; 98WO-US25247.  
 XX PR 25-NOV-1997; 97US-0066517.  
 XX PA (GEO) GEN HOSPITAL CORP.  
 XX PI Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;  
 XX PI Rahme LG, Tan M, Tsongalis J;  
 XX DR WPI; 1999-357851/30.  
 XX PT Virulence factors useful in developing disease treatments  
 XX PS Claim 12; Fig 9; 228pp; English.  
 XX CC The present sequence represents a Pseudomonas aeruginosa polypeptide  
 CC sequence. P. aeruginosa is an opportunistic human pathogen present in  
 CC soil water and plants. The specification describes virulence polypeptides  
 CC and nucleic acid sequence encoding such polypeptides. These sequences  
 CC can be used to identify a compound which is capable of decreasing the  
 CC expression of a pathogenic virulence factor. Compounds that inhibit  
 CC the expression or activity of virulence factor polypeptides can be  
 CC used to treat pathogenic infections, especially where the infection  
 CC is a P. aeruginosa infection.  
 CC note: the sequences given in the specification were poorly legible, and

CC in some instances assumptions were made as to the identity of the  
 CC residue; it is therefore possible that the sequence given below is  
 CC not entirely correct.

XX SQ Sequence 344 AA;

Query Match 6.9%; Score 100.5; DB 20; Length 344;  
 Best Local Similarity 22.7%; Pred. No. 0.019;  
 Matches 70; Conservative 47; Mismatches 128; Indels 63; Gaps 14;  
 QY 6 LDTLLSIIDEGSFEGASLALISPSAVSQRVKALEHHVGRVLSRT-OPAKATEAGEVLV 64  
 Db 9 vnmfqlvqasgsaarlkrkshsavsnleidlclvelvrrdgykvepteqalrli 68  
 QY 65 QAARWVLLQAEKTKAQLSGRLAEIPLTTAINADS-----LSTWFFPVFNEVASWGGA 116  
 Db 69 pymrsllnyq-----qlig-----diafnlnkprnlrvldtaippsfcd----- 109  
 QY 117 TLTLEDEAHTLSLLRR--GDVLTREANPVGCEVVE-----LGTMRHL 162  
 Db 110 tvssvldldfmmyslirtspadlatikqdaeidaitideelkisirfnqcvlgytkaf 169  
 QY 163 AIATPS---LRDAYWVDGKLDMAAMPV-LRFGPKDVLQDRDLGRVDGPVGRRRYSIVP 217  
 Db 170 vvahpqhplcmashlsiaslanyrqisrgsrqghsmll-----rpvs-dkvlfve 219  
 QY 218 SAEGFGEAIRRGGLGWLGPETQAPMLKAGEVILLDEI---PIDTPMY-WQWRLES-R 271  
 Db 220 nfdcmrlrveagvgwgiaphyfvfeerlrrngtlaviselypggldtkvycyntaleser 279  
 QY 272 SLARLTDA 279  
 Db 280 sfrifles 287

RESULT 9  
 AAB18640  
 ID AAB18640 standard; Protein; 1346 AA.  
 XX AC AAB18640;  
 XX DT 22-JAN-2001 (first entry)  
 XX DE Amino acid sequence of narbonolide synthase subunit 4 (PICATV).  
 XX KW Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;  
 XX KW antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;  
 XX KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;  
 XX KW picromycin biosynthesis.  
 XX OS Streptomyces venezuelae.  
 XX PN US6117659-A.  
 XX PD 12-SEP-2000.  
 XX PF 27-MAY-1999; 99US-0320878.  
 XX PR 28-MAY-1998; 98US-0087080.  
 XX PR 22-SEP-1998; 98US-0100880.  
 XX PR 08-FEB-1999; 99US-0119139.  
 XX PR 20-MAY-1999; 99US-0134990.  
 XX PR 30-APR-1997; 97US-0846247.  
 XX PR 06-MAY-1998; 98US-0073538.  
 XX PR 28-AUG-1998; 98US-0141908.  
 XX PA (KOSA-) KOSAN BIOSCIENCES INC.  
 XX PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;  
 XX WPI; 2000-610844/58.  
 XX

PT New recombinant pick hydroxylase gene of Streptomyces venezuelae useful  
 PT for converting ketolides to antibiotics and as antibiotics and  
 PT intermediates in the synthesis of compounds with pharmaceutical value  
 XX  
 XX  
 PS Disclosure; Columns 13-14; 117pp; English.  
 XX  
 CC The present sequence represents a narbonolide synthase subunit 4  
 CC (PikAIV). The nucleotide sequence encoding it is used in the course of  
 CC the invention. The specification describes a recombinant DNA compound  
 CC expressing recombinant polyketide synthase genes in host cells for the  
 CC production of narbonolide, narbonolide derivatives and polyketides that  
 CC are useful as antibiotics and as intermediates in the synthesis of  
 CC compounds with pharmaceutical value. The DNA compounds may also encode  
 CC a C12-hydroxylase (picK), desosamine biosynthesis and desosaminyl  
 CC transferase enzymes (useful for conversion of ketolides to antibiotics),  
 CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).  
 CC These compounds are also useful for increasing the antibiotic activity  
 CC of a compound relative to the unhydroxylated compound. The recombinant  
 CC host cells are useful as genetic systems that allow rapid engineering  
 CC of the narbonolide polyketide synthase. These would be valuable for  
 CC creating novel ketolide analogs for pharmaceutical applications.  
 XX  
 SQ Sequence 1346 AA;

Query Match 6.8%; Score 99.5; DB 21; Length 1346;  
 Best Local Similarity 20.9%; Pred. No. 0.18;  
 Matches 67; Conservative 35; Mismatches 107; Indels 111; Gaps 14;  
 QY 19 EGASLALISPSAVSORVKALEHHVGRVLSRTQPAKATEAG-----EVLVQAR----- 68  
 Db 688 kggmisalalseeatrgrienlh---glsiaavngptatvsgdptqigelaqaceadgir 744  
 QY 69 -KVVLLQ-----AETRAQLSGRLAEIPLTIAINADSLSTW----- 102  
 Db 745 ariipdvashshavtieneladvlagispqtpqpfstle---gtwitepalddy 800  
 QY 103 -----FPVFNVEVASWGAT-----ITLRLDEAHTLSLRRGDLGAVT 142  
 Db 801 wrynhrvgfapavetlatdefthfievshahpvtlmtlpdkvtglatlrred----- 854  
 QY 143 REANPVAGCEVVELGTMRHLAIATPSLRDAYMYDGLDWAAM-----PVLRFPGKDV 194  
 Db 855 -----ggqhrli---ttslaeawanglaldwasllpatgalspavpdptiya 897  
 QY 195 LQDRDLGDRVDPVGRRRRSIVPSAEGFGEAIRRGLGWGLLPE-----TQAAPML 244  
 Db 898 fghrsywisagpg-geaphtasgreavaet---glawpggaedldeegrissavlammvr 953  
 QY 245 KAGEVILDD---EIPIDTPM 261  
 Db 954 gaasvlrcdspeevpvdrl 973

## RESULT 10

AAV77195  
 ID AAY77195 standard; Protein; 1346 AA.

XX  
 AC AAY77195;

XX  
 DT 05-JUN-2000 (first entry)

XX  
 DE S. venezuelae macrolide biosynthetic enzyme PikAIV, SEQ ID NO:37.

XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,  
 KW chronic obstructive pulmonary disease; respiratory inflammation;  
 KW hypercholesterolaemia; crop protection agent.  
 XX  
 OS Streptomyces venezuelae ATCC15439.

PN WC200000620-A2.  
 XX  
 PD 06-JAN-2000.  
 XX  
 PF 25-JUN-1999; 99WO-US14398.  
 XX  
 PR 26-JUN-1998; 98US-0105537.  
 XX  
 PA (MINU ) UNIV MINNESOTA.  
 XX  
 PI Sherman DH, Liu H, Xue Y, Zhao L;  
 XX WPI; 2000-160679/14.  
 DR N-PSDB; AA287300.  
 XX  
 XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.  
 PT synthesis of methymycin and pikromycin -  
 PS Claim 19; Page 424-428; 438pp; English.  
 XX  
 CC The invention relates to an isolated and purified nucleic acid segment  
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its  
 CC biologically active variant, where the nucleic acid sequence is not  
 CC derived from the eryC gene cluster of Saccharopolyspora erythraea or  
 CC Streptomyces antibioticus. The invention also relates to a macrolide  
 CC biosynthetic gene cluster, or fragments thereof. The macrolide  
 CC biosynthetic gene cluster encodes proteins which synthesise methymycin,  
 CC pikromycin, neomethymycin, narbomycin or a combination of these  
 CC compounds. Recombinant or augmented cells comprising the desosamine  
 CC and/or macrolide biosynthetic gene clusters are useful for the  
 CC production of biologically active macrolides. The macrolide biosynthetic  
 CC proteins are useful for synthesis of methymycin, pikromycin,  
 CC neomethymycin and narbomycin. The alternative termination of polyketide  
 CC synthesis may be useful to prepare novel antibiotics and  
 CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the  
 CC recombinant host cells are useful as biopolymers, e.g., in packaging or  
 CC biomedical applications, to engineer PHA monomer synthases or to prepare  
 CC biologically active agents, such as chemotherapeutics,  
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
 CC disease as well as other diseases involving respiratory inflammation,  
 CC cholesterol-lowering agents or macrolide-based antibiotics which are  
 CC active against a variety of organisms, e.g., bacteria, including  
 CC multi-drug resistant pneumococci and other respiratory pathogens, as well  
 CC as viral parasitic pathogens, or as crop protection agents (e.g.,  
 CC fungicides or insecticides) via expression of polyketides in plants.  
 CC Sequences AAY77190-Y77197 represent macrolide biosynthetic enzymes from  
 CC Streptomyces venezuelae ATCC 15439, which are encoded by sequences  
 CC AAZ87295-Z87302.  
 XX  
 SQ Sequence 1346 AA;

Query Match 6.8%; Score 99.5; DB 21; Length 1346;  
 Best Local Similarity 20.9%; Pred. No. 0.18;  
 Matches 67; Conservative 35; Mismatches 107; Indels 111; Gaps 14;

QY 19 EGASLALISPSAVSORVKALEHHVGRVLSRTQPAKATEAG-----EVLVQAR----- 68  
 Db 688 kggmisalalseeatrgrienlh---glsiaavngptatvsgdptqigelaqaceadgir 744  
 QY 69 -KVVLLQ-----AETRAQLSGRLAEIPLTIAINADSLSTW----- 102  
 Db 745 ariipdvashshavtieneladvlagispqtpqpfstle---gtwitepalddy 800  
 QY 103 -----FPVFNVEVASWGAT-----ITLRLDEAHTLSLRRGDLGAVT 142  
 Db 801 wrynhrvgfapavetlatdefthfievshahpvtlmtlpdkvtglatlrred----- 854  
 QY 143 REANPVAGCEVVELGTMRHLAIATPSLRDAYMYDGLDWAAM-----PVLRFPGKDV 194  
 Db 855 -----ggqhrli---ttslaeawanglaldwasllpatgalspavpdptiya 897  
 QY 195 LQDRDLGDRVDPVGRRRRSIVPSAEGFGEAIRRGLGWGLLPE-----TQAAPML 244

Db 898 fghrsywisagp-geapahtasgreavaet---glawpggaedideegrrsavlamvmr 953  
 QY 245 KAGEVILLD---EIPIDTPM 261  
 Db 954 qaasvlrcdspeevpdrpl 973

RESULT 11

AAAY67204  
 ID AAY67204 standard; protein; 1346 AA.  
 AC AAY67204;  
 XX  
 XX 23-MAR-2000 (first entry)  
 DE Narbonolide synthase subunit 4 (PICAIV) protein sequence.  
 XX Narbonolide polyketide synthase; PKS; narbonolide synthase subunit 4;  
 KW PICAIV; antibiotic production; narbomycin; picromycin; ketolide.  
 XX Streptomyces venezuelae.

OS  
 XX WO9961599-A2.  
 PN  
 XX 02-DEC-1999.  
 PD  
 XX 27-MAY-1999; 99WO-US11814.  
 PF  
 XX 28-MAY-1998; 98US-0087080.  
 PR 28-AUG-1998; 98US-0141908.  
 PR 22-SEP-1998; 98US-0100880.  
 PR 08-FEB-1999; 99US-0119139.  
 XX  
 XX (KOSA-) KOSAN BIOSCIENCES INC.

XX Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;  
 PI WPI; 2000-072618/06.  
 DR N-PSDB; AAZ56001.  
 DR  
 XX New recombinant DNA encoding a domain of narbonolide polyketide  
 PT synthase, for production of ketolide antibiotics -  
 XX Example 2; Page 14-15; 98pp; English.

XX This is the Streptomyces venezuelae narbonolide synthase subunit 4,  
 CC PICAIV protein sequence. The invention relates to recombinant DNA  
 CC containing a coding sequence for a narbonolide polyketide synthase  
 CC (PKS). Polyketides are compounds synthesised from 2-carbon units through  
 CC a series of condensations and subsequent modifications. Modular PKSs are  
 CC responsible for the production of many antibiotics including picromycin.  
 CC The narbonolide PKS consists of a loading module, six extender modules,  
 CC and two thioester domains. Four proteins make up the narbonolide PKS  
 CC (PICAIV, PICAII, PICAIII and PICAIV). PICAIV includes the loading module  
 CC and extender modules 1 and 2, PICAII includes extender modules 3 and 4,  
 CC PICAIII includes extender module 5 and PICAIV includes extender module 6  
 CC and a type II thioesterase domain. The second type II thioesterase  
 CC domain is found on the PICB protein. The nucleotide sequences encoding  
 CC all of these proteins can be isolated in recombinant form from the  
 CC recombinant cosmid pK05023-27 (see AAZ56001). Narbonolide is  
 CC desaminylated in S. venezuelae to yield narbomycin, the desosamine  
 CC transferase enzyme is required for this conversion, and the desosamine  
 CC biosynthetic genes are also found in cosmid pK05023-27. The recombinant  
 CC DNA of the invention is used to express, in transformed cells,  
 CC narbonolide (or its derivatives) or other ketolides (particularly  
 CC hybrids), which may then be converted (e.g. by other enzymes  
 CC recombinantly expressed in the same hosts) to polyketide antibiotics or  
 CC their intermediates. The antibiotics are useful in human or veterinary  
 CC medicine.

XX Sequence 1346 AA;

Query Match 6.8%; Score 99.5; DB 21; Length 1346;  
 Best Local Similarity 20.9%; Pred No. 0.18;  
 Matches 67; Conservative 35; Mismatches 107; Indels 111; Gaps 14;  
 QY 19 EGASIALSISPAVSQRVKALEHHVGRVLVSRTPAKATEAG-----EVLVQAAR----- 68  
 Db kggmislalseeatqrriehlh---glslaaavngptatvsgdptqigelaqaceadgir 744  
 QY 69 -KMWLLQ-----AETKAQLSGRLAEPLTIAINADSLSTW----- 102  
 Db 745 arlipdyashahvetieneladvgispdtpqpvffstle---gtwitepaldggy 800  
 QY 103 -----PPVFNEVASWGGAT-----LTLRLDEAHTLSLLRRGDVLGAVT 142  
 Db 801 wynlrhrvgfapavetlatdegfthfievshahpvtmtltpdkvtglatlrred----- 854  
 QY 143 REANPVAGEVVELGTMRLHAIATPSLRDAYMVDGKLDWAAM-----PVLRFPGKDV 194  
 Db 855 -----ggqhrh---ttslaeawangialdwasilpatgalispavpdilpya 897  
 QY 195 LODRLDGRVDGPVGRRRVSIIVPSAEGFGEATRRGLGMLLPE-----TQAAPML 244  
 Db 898 fghrsywisagp-geapahtasgreavaet---glawpggaedideegrrsavlamvmr 953  
 QY 245 KAGEVILLD---EIPIDTPM 261  
 Db 954 qaasvlrcdspeevpdrpl 973

RESULT 12

AAAY77180  
 ID AAY77180 standard; Protein; 12199 AA.  
 XX  
 AC AAY77180;  
 XX  
 XX 05-JUN-2000 (first entry)  
 DT  
 XX S. venezuelae pik gene cluster-encoded protein, SEQ ID NO:6.  
 DE  
 XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,  
 KW chronic obstructive pulmonary disease; respiratory inflammation;  
 KW hypercholesterolaemia; crop protection agent.  
 XX Streptomyces venezuelae ATCC15439.  
 OS  
 XX WO200000620-A2.  
 PN  
 XX 06-JAN-2000.  
 PD  
 XX 25-JUN-1999; 99WO-US14398.  
 PF  
 XX 26-JUN-1998; 98US-0105537.  
 PR  
 XX (MINU ) UNIV MINNESOTA.  
 PA  
 XX Sherman DH, Liu H, Xue Y, Zhao L;  
 PI WPI; 2000-160679/14.  
 XX N-PSDB; AAZ87285.  
 DR  
 XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.  
 PT synthesis of methymycin and pikromycin -  
 XX Claim 19; Page 315-353; 438pp; English.  
 PS The invention relates to an isolated and purified nucleic acid segment  
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its  
 CC biologically active variant, where the nucleic acid sequence is not  
 CC derived from the eryC gene cluster of Saccharopolyspora erythraea or  
 CC Streptomyces antibioticus. The invention also relates to a macrolide

CC biosynthetic gene cluster, or fragments thereof. The macrolide  
 CC biosynthetic gene cluster encodes proteins which synthesise methymycin,  
 CC pikromycin, neomethymycin, narbomycin or a combination of these  
 CC compounds. Recombinant or augmented cells comprising the desosamine  
 CC and/or macrolide biosynthetic gene clusters are useful for the production  
 CC of biologically active macrolides. The macrolide biosynthetic proteins  
 CC are useful for synthesis of methymycin, pikromycin, neomethymycin and  
 CC narbomycin. The alternative termination of polyketide synthesis may be  
 CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)  
 CC monomers. The compounds produced by the recombinant host cells are useful  
 CC as biopolymers, e.g., in packaging or biomedical applications, to  
 CC engineer PHA monomer synthases or to prepare biologically active agents,  
 CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,  
 CC chronic obstructive pulmonary disease as well as other diseases involving  
 CC respiratory inflammation, cholesterol-lowering agents or macrolide-based  
 CC antibiotics which are active against a variety of organisms, e.g.,  
 CC bacteria, including multi-drug resistant pneumococci and other  
 CC respiratory pathogens, as well as viral parasitic pathogens, or as crop  
 CC protection agents (e.g., fungicides or insecticides) via expression of  
 CC polyketides in plants. The present sequence represents a protein  
 CC encoded by the macrolide biosynthetic gene cluster (pik) from  
 CC Streptomyces venezuelae ATCC 15439.  
 XX  
 SQ Sequence 12199 AA;

Query Match 6.8%; Score 99.5; DB 21; Length 12199;  
 Best Local Similarity 20.9%; Pred. No. 4.6;  
 Matches 67; Conservative 35; Mismatches 107; Indels 111; Gaps 14;  
 QY 19 EGASLALISPSAVSRVKALEHHVGRVLSRTPQAKATEAG-----EVLVQAAR----- 68  
 Db 11260 kgmislalseeatrqrienlh---glisiaavngptatvsvgdpqtqigelaqaaceadgir 11316  
 QY 69 -KVVLLQ-----AETKAOLSGRLAEIPLTIAINADSLSTW----- 102  
 Db 11317 ariipdvashshvieteneladvlaglspqtpqvffstle---gtwitepalddgy 11372  
 QY 103 -----FPPVFNESVWGKAT-----LFLRLEDEAHTLSLRRGDLVGAVT 142  
 Db 11373 wrynrlhrvgfapavetlatdegthfievshahpvlmtipdkvtglatirred----- 11426  
 QY 143 REANPVAGCEVVELGTRHLAIATPSLRDAYWVDGKLDWAM-----PVLRFPGKDV 194  
 Db 11427 -----ggqhr1---ttsiaewanglaidwasllpatgalspavpdlptya 11469  
 QY 195 LQDRDLGRVDGPGVRRRSVTSVSAEGFGAIRRGGLWGLLPE-----TQAAPML 244  
 Db 11470 fhrsywispagp-geapahtasgreavaet---glawgpgaedldeegrrsavlanvmr 11525  
 QY 245 KAGEVILLD---EIPIDTPM 261  
 Db 11526 qaasvlrcdspeevpvdrl 11545

RESULT 13  
 AAY74591  
 ID AAY74591 standard; Protein: 303 AA.  
 XX  
 AC AAY74591;  
 XX  
 DT 21-MAR-2000 (first entry)  
 XX  
 DE Neisseria gonorrhoeae ORF 158 protein sequence SEQ ID NO:656.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KW antibacterial; gene therapy.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 PN WO9957280-A2.  
 XX

PD 11-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-US09346.  
 XX  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 09-OCT-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 25-FEB-1999; 98US-0103796.  
 XX  
 XX (CHIR) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX  
 DR WPI; 2000-062150/05.  
 DR N-PSDB; AAZ53353.  
 XX  
 PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 XX  
 PS Claim 2; Page 452; 1453pp; English.  
 XX  
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941  
 CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX  
 SQ Sequence 303 AA;  
 Query Match 6.7%; Score 98; DB 21; Length 303;  
 Best Local Similarity 27.6%; Pred. No. 0.029;  
 Matches 43; Conservative 28; Mismatches 55; Indels 30; Gaps 7;  
 QY 2 NPQLDTLLSIIDEGSPEGASLALISPSAVSRVKALEHHVGRVLSR-TQPAKATEAG 60  
 Db 4 nseeltvfvqvesgsfrraaeqlemansavsrivkrleeklgvnltrtrglnlteeg 63  
 QY 61 EVLVQARAKVLLQAEETKAOLSGRLAEIPLTIAINADSLSTWPPV-----FNE- 109  
 Db 64 aqyfriragilqemaetaemula-vhevpggv-irvds---mpmvlhllaplaakfner 118  
 QY 110 -----VASWGGATLTLRLEDEAHTLSLRRGDV 137  
 Db 119 yphirlsvssegynlierkvdia-----lragel 149  
 RESULT 14  
 AAW52846  
 ID AAW52846 standard; Protein: 5069 AA.  
 XX  
 AC AAW52846;  
 XX  
 DT 24-JUL-1998 (first entry)  
 XX  
 DE A. mediterranei rifamycin synthesis gene cluster fragment protein B.  
 KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;  
 XX

KW polyketide synthase; actinomycete; ansamycin.

OS Amycolatopsis mediterranei.

PN WO9807868-A1.

XX 26-FEB-1998.

XX 18-AUG-1997; 97WO-EP04495.

XX 20-AUG-1996; 96EP-0810551.

XX (NOVS ) NOVARTIS AG.

XX Engel N, Schupp T, Toupet C;

XX WPI: 1998-169172/15.

DR N-PSDB; AAV211187.

XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used  
PT to produce rifamycin and rifamycin analogues

XX Claim 6; Page 126-151; 205pp; English.

CC The present sequence represents a Amycolatopsis mediterranei rifamycin  
CC synthesis gene cluster ORF B protein from the present invention. The  
CC DNA fragment comprises a B region involved directly or indirectly  
CC in the gene cluster responsible for rifamycin synthesis, including  
CC the adjacent DNA regions to the right and left which, by reason of  
CC their function in connection with rifamycin biosynthesis, qualify  
CC as constituents of this rifamycin gene cluster, and functional  
CC fragments, derivatives or constituents of these. The Amycolatopsis  
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used  
CC for producing rifamycin, rifamycin analogues or precursors. It can also  
CC be used for inactivating or modifying genes involved in ansamycin or  
CC rifamycin biosynthesis. The DNA can be used for constructing mutant  
CC actinomycetes strains from which the natural rifamycin or ansamycin  
CC biosynthesis gene cluster has been partly or completely deleted. The  
CC DNA fragment can be used for assembling a library of polyketide  
CC synthases, which can be used for assembling a library of polyketides.  
CC A hybridisation probe of the invention can be used for identifying DNA  
CC fragments involved in the biosynthesis of ansamycins.

XX Sequence 5069 AA;

Query Match 6.5%; Score 95.5; DB 19; Length 5069;

Best Local Similarity 23.7%; Pred. No. 3.4; Mismatches 79; Gaps 15;  
Matches 71; Conservative 45;

QY 21 ASLALISPSAVSQRV-----KALEHHVGRVLSVRTQPAKATEAGEVLVQAAARKWVLLQA 75

Db 3919 asfamvvglaavqsgvgrpdavghsqgei-----aaacvsgalslqdaakvvalrs 3971

QY 76 EYKA-QLSGR--LAELPLTIANADSLSTWTFPP-----VFNEVASWGGATLRLDED 124

Db 3972 qalatrlagrgma-----svalseedatawlqpdwrvqvaavnspsvviageaqalde 4027

QY 125 EAHTLS---LLRRGDV-LGAVTRE-----ANPVACEVVELGTMRHLAATPSLRD 171

Db 4028 vvdalsggvrvrvavdygshntqveatedllaetlagiesqapkvpfystligdwird 4087

QY 172 AYMDVKLDWAAMPVLRFGPKDVLQDRDLGRVDGPGVGRRRRSIVPSAEGFGEAIRRGL- 230

Db 4088 agivdgyggy-wy-----rnlrnqv-----gfgpavaelvr 4115

QY 231 -CWGLLPETQAAPMLKAGEVILLDELIPDTPMWNORWRLSRLARLTDAVDAIEGL 288

Db 4116 qdnhgvfvevsahpvl-----vqplsel-s-ddavvtgslrredgglrlrltsmaelyvqgv 4169

RESULT 15  
c AAR25588

ID AAR25588 standard; Protein; 298 AA.

XX AAR25588;

XX 08-JAN-1993 (first entry)

XX Truncated irgB.

XX Virulence factor; intestine; multi-valent vaccine.

XX Vibrio cholerae.

XX WO9211354-A.

XX 09-JUL-1992.

XX 18-DEC-1991; 91WO-US09592.

XX 18-DEC-1990; 90US-0629102.

XX (GEHO ) GEN HOSPITAL CORP.

XX (HARD ) HARVARD COLLEGE.

XX Calderwood SB, Goldberg MB, Mekalanos JJ;

XX WPI: 1992-250077/30.

XX N-PSDB; AAQ26543.

XX Cholera vaccine comprising live attenuated Vibrio cholerae cells  
PT - contains mutations which inhibit expression of a functional  
PT irgA gene product, vaccine has reduced virulence, decreasing side  
PT effects

XX Disclosure; Fig 13; 75pp; English.

XX The functional irgB gene encodes a protein which acts as a positive  
CC regulator of irgA. irgA functions as a V. cholerae virulence factor.  
CC The sequence given shows a truncated irgB. This truncation of the  
CC sequence inhibits expression of a functional irgA gene product in the  
CC cell. This sequence can be used to transform V. cholerae cells which  
CC can then be used to induce immunity to cholera in a mammal. These  
CC transformed cells can colonize the intestines of inoculated animals and  
CC may be combined with vaccines targeted at other illnesses to make a  
CC single multi-valent vaccine. The transformed cells have reduced  
CC virulence such that the inoculated animal is less likely to develop  
CC side effects such as diarrhoea and fever.

XX Sequence 298 AA;

Query Match 6.4%; Score 93.5; DB 13; Length 298;

Best Local Similarity 23.0%; Pred. No. 0.087; Mismatches 33; Conservative 33; Indels 71; Gaps 11;

QY 17 SFEGLASLISPSAVSQRVKALEHHVGRVLSVRT-QPAKATEAGEV-----LVQAAAR 68

Db 18 sitaakaaleqkqstlsrrlaqleedlqgslImrggnriltkagevfayvseqillelan 77

QY 69 KVVLLQAEKTKAOLSGRLAEIPLTIANADSLSTWTFPPVFNEVASWGGATLTLRL----- 122

Db 78 Ksqaalqelnngvtge-----ltlvvhpnlirgwlsvqvidefmq-ghstikirlisqfgh 131

QY 123 EDEAHTLSLLRRGDVLGAVTREANPNVAGCEVVELGTMRHLAATPSL-----RDAYMVDGKL 179

Db 132 sdevfedpdl-----iwiehaapmgyrkerlygyrlyatyaspkylahrd----- 175

QY 180 DMAAMPVLRFGPKDVLQDRDLGRVDGPGVGRRRRSIVPSAEGFG----- 223

Db 176 ----kpt---hprelihnpwidf-----iacrraelelhphfsgysipalesrlqsdnl 223

QY 224 ----EAIIRGLGWGLLP 236

Db 224 amqadaakgrgigllp 240

Search completed: September 13, 2001, 17:51:47  
Job time: 1118 sec





Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	136.5	9.3	276	1	US-07-952-817-27	Sequence 27, Appl	
2	118.5	8.1	278	4	US-09-036-987A-22	Sequence 22, Appl	
3	99.5	6.8	1346	3	US-09-320-878-4	Sequence 4, Appl	
4	96	6.6	251	1	US-07-952-817-28	Sequence 28, Appl	
5	92.5	6.3	4150	4	US-09-428-517-2	Sequence 2, Appl	
6	89	6.1	290	2	US-08-614-686A-1	Sequence 1, Appl	
7	89	6.1	840	2	US-08-467-822-25	Sequence 25, Appl	
8	89	6.1	840	4	US-08-432-697-25	Sequence 25, Appl	
9	87.5	6.0	433	4	US-09-385-028-10	Sequence 10, Appl	
10	87.5	6.0	3798	3	US-09-335-409-6	Sequence 6, Appl	
11	86.5	5.9	3031	1	US-07-689-008-2	Sequence 2, Appl	
12	85.5	5.9	291	1	US-07-952-817-26	Sequence 26, Appl	
13	85	5.8	1068	1	US-08-537-210A-2	Sequence 2, Appl	
14	85	5.8	1068	4	US-09-113-825-2	Sequence 2, Appl	
15	85	5.8	2556	1	US-08-185-432-17	Sequence 17, Appl	
16	84.5	5.8	3739	3	US-09-320-878-2	Sequence 2, Appl	
17	84	5.8	829	4	US-09-413-814-105	Sequence 105, Appl	
18	84	5.8	1548	1	US-08-463-092B-7	Sequence 7, Appl	
19	84	5.8	1548	2	US-08-460-907B-7	Sequence 7, Appl	
20	83.5	5.7	591	4	US-09-413-814-76	Sequence 76, Appl	
21	83.5	5.7	3816	4	US-09-428-517-3	Sequence 3, Appl	
22	82.5	5.7	2152	4	US-09-036-987A-3	Sequence 3, Appl	
23	82.5	5.7	4472	2	US-08-804-227C-2	Sequence 2, Appl	
24	81.5	5.6	503	3	US-08-911-853-7	Sequence 7, Appl	
25	81.5	5.6	503	4	US-09-479-409-7	Sequence 7, Appl	
26	81.5	5.6	2509	1	US-08-469-005A-10	Sequence 10, Appl	
27	81.5	5.6	4545	2	US-08-804-227C-14	Sequence 14, Appl	

```

Qy 60 GEVLVQAARKMV--LLQAETKAQUSGLAEIPIUTIAINADSLSTWFPFVFNVASWGAT 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 GAMLITTEGRATIGALAAAERTQRAA--TSQPSLVLAAKAGACAGCELLAKLLDAYAEPGAV 119
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 118 LTLKLEDAHTLSLLR--RGDVLGAVTRENPAVAGEVVELGTMRHLATATPSLRDAYV 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 AVDLLLCESPQKTLHDGRADV--ALLHQPDPTAELDIEILNTEQGVAILPTS----- 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 176 DGKLDWAAMPVLRFGPKDVLQDRDLDGRVDGPGVGRNRVSTVPSAEG-----FGEAIR 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 ---HPLASEPHVRMDVSSLPDPL-ARWPGDG-----YIPDGPGEVNRNQTLEQMIA 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 228 RGLGWGLLPETAAPMLKA-GEVILLDEIPIDTPMYQWRRLSESLARL 276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 LGRITVMPSSSRNVNLEGLAAVPVLDADPNVTVTIAWPP--HSPSRLAAGI 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 3  
US-09-320-878-4  
: Sequence 4, Application US/09320878A  
: Patent No. 6117659  
: GENERAL INFORMATION:  
: APPLICANT: ASHLEY, Gary  
: APPLICANT: BETLACH, Melanie C.  
: APPLICANT: BETLACH, Mary C.  
: APPLICANT: MCDANIEL, Robert  
: APPLICANT: TANG, Li  
: TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
: FILE REFERENCE: 300622002120  
: CURRENT APPLICATION NUMBER: US/09/320,878A  
: CURRENT FILING DATE: 1999-05-27

[illegible]

```

Query Match      6.6%; Score 96; DB 1; Length 251;
Best Local Similarity 21.7%; Pred. No. 0.0093;
Matches 59; Conservative 44; Mismatches 101; Indels 68; Gaps 10;

QY 10 LSIIDEGFEGASLALISPSAYSORVKALEHHVGRVLVS-TPAKATEAGEVLVQAAR 68
      : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 10 VTVVEQSISKAERKLCIAQPPLSRQIQKEELGICIQLFERGFPAKVTEAGMFFYQHA- 68
      : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 69 KVVLLQAEQAQISGRUAETPLTIAIN--ADSLSTWPPVPFNEVASMGGATLTLRLLEDEA 126
      : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 69 VQILTHTAQASSNAKRATVSYOTLRIGVSSLLYGLLPEII----- 109
      : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 127 HTLSLLRRGDVLGAVTREANPVACGEVVELGT-----MRHLAIAATPSLRD 171
      : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 110 -----YLFQOQNPETHIELIECGTKDQIINALKQGGKIDLGFGRLKITDPAIRR 156
      : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 172 AYVWDGKLDAWNP-----VLRFGPKDVLQDRDLGCRVDPGVRRRVSVPSNAGFGEAIRR 228
      : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 157 IVLHKEQGLKIAIHKHHNLNOFAATGV---HLSQIIDPEMLLYPVSQKPNFATIQSLFT 212
      : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 229 GLGWGGLP-----ETQ-AAPMLKAGEVILL 252

```

```

; CITY: Roslyn
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11576
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/614,686A
; FILING DATE: MARCH 12, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GR195110930
; FILING DATE: 24 MARCH 1995
; APPLICATION NUMBER: GR19514056.7
; FILING DATE: 13 APRIL 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Collard, Allison C.
; REGISTRATION NUMBER: 22,532
; REFERENCE/DOCKET NUMBER: BOCK ET AL. -4
; ATTORNEY/AGENT INFORMATION:
; NAME: Freedman, Edward R.
; REGISTRATION NUMBER: 26,048
; REFERENCE/DOCKET NUMBER: BOCK ET AL. -4
; ATTORNEY/AGENT INFORMATION:
; NAME: Keusey, Edwin H.
; REGISTRATION NUMBER: 34,361
; REFERENCE/DOCKET NUMBER: BOCK ET AL. -4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 365-9802
; TELEFAX: (516) 365-9805
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 290 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Klebsiella terrigena
; STRAIN: DSM2867
; US-08-614-686A-1

```

Query Match 6.1%; Score 89; DB 2; Length 290;  
 Best Local Similarity 23.6%; Pred. No. 0.071;  
 Matches 74; Conservative 56; Mismatches 105; Indels 78; Gaps 20;

```

QY 10 LSTIDSGFEGASLALISPSAVSQRVKALEHHVGRVLSR-TQPAKATEAGE-VLVQAA 67
Db 10 VAVAEARNTRAHDGIGSQPLSQIQRLEREIGTGPLRLTRGVELTEAGESFYVDAC 69
QY 68 RKMVLQAA---ETKA---OLSGRLAE-IPLTAINADSLSTW-----PPPVFNEVASWG 114
Db 70 QIALSDAALEKTKGIARGMNGSLVPGITSSAAFSHQIESLLYQFOQRYPAVALRQVEGN 129
QY 115 GATLT-----LRL---EDEAHTLSLLRRCVDVLGAVTREANPVAGCEVVELGTM 159
Db 130 MATLHMALGEAEIDTAFLVPCSSKAFNLRIIAEPMVIALHR-SHPLSG----- 179
QY 160 RHLATATPSLRDAYMVDGLDWAAMPVLRFGPKDV--LQDRLDGL--RVDPGVGR-RRVS 214
Db 180 -ESALSQAQSD-----AVPVI-FPEVAPGLYEQVYDGCRRAGVDMRSARQSS 226
QY 215 IYPSAGGFEATRGWGGLLPETQAA-----PMLKAGEVILLDIPIDTMYWQR 265
Db 227 QISSISIMVDA---GFGFALVPQSMTCICLPNVWTHPLQDAS---LKTEIAI-----AWRR 276
QY 266 WRLESRLARLTD 278
Db 277 FE-RSETVKRFLE 288

```

RESULT 7

```

US-08-467-822-25
; Sequence 25, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,822
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 840 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-467-822-25

```

Query Match 6.1%; Score 89; DB 2; Length 840;  
 Best Local Similarity 22.1%; Pred. No. 0.36;  
 Matches 59; Conservative 30; Mismatches 114; Indels 64; Gaps 10;

```

QY 34 QRVKALEHHVGRVLSRTPAKATEAGEVLVQAAKRWLLQAEKTAQLSGRLAEIPLTTA 93
Db 69 QVLPAVPHLLNAVQVENTEP-----DGTKLVTVDHPISRENGELQALFGLSLPVP--- 119
QY 94 INADSLSTWFPVFNVEVASWGATLTLRLEDEAHTLSLLRRCVDVLGAVTREANPVAGCV 153
Db 120 ----SLDK-----FATETKEDNRIPGEILCEDECLTINIGRKAVILKVKTSKGRDP----- 164
QY 154 VELGTMRHLAIATPSL-----RDAYMVDGLDWAAMPVLRFGPKDVLQDR--DLQGR--- 203
Db 165 IOGVSHYHFIEVNPVLTFRRKAYGM--RLNTAAGTAVRFEFGDCKSVTLVSIEGNKVR 222

```

Qy	204	-----VGPVGRRRVSVPSA---EGGGEAIRCLGWLGLLPE-----	237
Db	223	GGNAIADGPVNETLEAAHMAVRSGEGHEEKDAPEGFTKEDPNCNSFNTIHRKEYANK	282
Qy	238	--TQAAPMLKAGEVILLDEIPIDTPMY	262
Db	283	YGPTTGDKIRLGDTNLLAEIEKQVALLY	309

```

RESULT 9
US-09-385-028-10
; Sequence 10, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen
; APPLICANT: Keamena A Aidoo
; APPLICANT: Ashish S. Paradkar
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
; TITLE OF INVENTION: Acid Biosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN PPLC
; STREET: The Jenner Bulding, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P57452US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202 638-6666
; TELEFAX: (202) 39305350
; TELEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-385-028-10

```

Query Match	6.08;	Score 87.5;	DB 4;	Length 433;
Best Local Similarity	33.7%;	pred. No. 0.19;		
Matches 29;	Conservative 16;	Mismatches 38;	Indels 3;	Gaps 2;
Qy	6	LDTLLSIDEGSPGASIALSISPSAVSQRYKALEHHVGRVLVSTQPAKATEAGEVLVQ	65	
Db	330	LHLQAVRHGSINRAAAVLSTISALPTRIHRLQSLGARLLRS--PRGTSLTGPTRQ	387	
Qy	66	AARKWVLLQAE--TKAQLSGRLAEIPL	90	
Db	388	FLQALYEAEFREAAALACRSVERPL	413	
RESULT	10			

US-09-335-409-6

; Sequence 6, Application US/09335409

; Patent No. 6121029

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/335,409

; CURRENT FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 6

; LENGTH: 3798

; TYPE: PRT

; ORGANISM: Sorangium cellulosum

US-09-335-409-6

Query Match 6.0%; Score 87.5; DB 3; Length 3798;

Best Local Similarity 22.2%; Pred. No. 5.3;

Matches 82; Conservative 29; Mismatches 117; Indels 141; Gaps 16;

QY 2 NPQLDTLLSIIDGSEFEGASLALSISPAVSQVKALE-----HH---42

Db 345 DPEIEALRAVVGPARADGARCVL-----GAVKTNLHLEGAAGVAGLIKATLSLHHERIP 400

QY 43 -----VGRVLVSRTPA-----KATEAGEVLVQAARKM 70

Db 401 RNLNFTLNPRIEGTALALATEPVWPTGRTREAGVSGMSTNAHVLEEAP---457

QY 71 VLLQATKKAQSLGRLEIPLTITAINADSLSTWPPVNEVASWGGATLRLLEDEAHTLS 130

Db 458 ---AVEPEAAAPRAAELEFVLSAKSAALD-----QAARLRDHLKXHV 498

QY 131 LRRRGDVLG--AVTREANVAGCEVVELGTMRH-LAIATPS---LRDAY-----173

Db 499 ELGLDVAFLATRSa-----MEHRLVAASSREALGALSAAGHTTPPG 545

QY 174 MYDGKLDNAAMPVLR-----GPKDVLQDRDLGRVGPVRRRSIVPSAEFGFGEAIRR 228

Db 546 AVRGRASGSGAPKVVFPFGQGSOWGMGRKL--MAEPEVFR-----AALEGCDRAIEA 597

QY 229 GLGWGLLPETQA-----APMLKAGEVILLDEIPDTPMYWQWRLESRLAR 275

Db 598 EAGWSLLGELSADAASQLGRIDVVPQVLFAMEVAL-----SALWRSWGVEPEAVVG 649

QY 276 LTDAAVVDAA 284

Db 650 HSMGEVAAA 658

RESULT 11

US-07-689-008-2

; Sequence 2, Application US/07689008

; Patent No. 5268274

; GENERAL INFORMATION:

; APPLICANT: Ben-Bassat, Arie

; APPLICANT: Calhoon, Roger D

; APPLICANT: Fear, Anna L

; APPLICANT: Gelfand, David H

; APPLICANT: Meade, James H

; APPLICANT: Tal, Ronny

; APPLICANT: Wong, Hing

; APPLICANT: Benzman, Moshe

; TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE

; TITLE OF INVENTION: EXPRESSION OF CELLULOSE SYNTHASE OPERON

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: McCutchen, Doyle, Brown & Enersen

; STREET: Three Embarcadero Center

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.24

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/689,008

; FILING DATE: 19910422

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 337,194

; FILING DATE: 12-APR-1989

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 496,236

; FILING DATE: 23-MAR-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Murphy, Lisabeth Feix

; REGISTRATION NUMBER: 31547

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 393-2000

; TELEFAX: (415) 393-2286

; TELEX: 340817 MACPAG SFO

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3031 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-07-689-008-2

Query Match 5.9%; Score 86.5; DB 1; Length 3031;

Best Local Similarity 21.6%; Pred. No. 4.9;

Matches 68; Conservative 36; Mismatches 94; Indels 117; Gaps 14;

QY 23 LALSTSPSAVSQVRKAL--EHHVGRVLSR-----TOPAKATEAGEVLVQ 65

Db 2120 LLAGLSPADYSPAIRSIAEMEIKDASRLSMVSNPVLIREALTQPDPTGARGVAVAD 2179

QY 66 AAR-----KWLLOAETKAQLSGRLEIPLTITAINAD---SLSTWPPVNEVAS--112

Db 2180 LFRQRGDWVHARMALRIASR-----TIDLSPDQRLSYATEYMKISNPVAAAR 2227

QY 113 -----WGGATLTLRLDEAHTLSLRRG-----DVLG-----AVTREAN 146

Db 2228 LLAPLGDGTSATGSALLPEQVOTLQQLRMGISVAGSDLLNQGDQAQAYDHLAPALQAD 2287

QY 147 PVAGCEVVELGTMRHLATIPSLRDAYWVDGKLDWAMPVLRFGPKD-----VL 195

Db 2288 PEA-----TSPKLALA--RLYNGHGKPGKALETDLAVLRHNPQDLDARQAQVAAVYN 2337

QY 196 QDRD-----LDGRVDGVPVGR--RVSVIPSAEFGF-----EAIR 227

Db 2338 SDHNSLATFLAMDGVQESPMADARAWLAMAVADQADGHGORTIEDLRLRDLRQVQEGTR 2397

QY 228 RGLGWGLLPETQAAP 242

Db 2398 AASGAGAAQEDALAP 2412

RESULT 12

US-07-952-817-26

; Sequence 26, Application US/07952817

; Patent No. 5356796

; GENERAL INFORMATION:

; APPLICANT: Keller, John W.

; TITLE OF INVENTION: A Repressor Protein and Gene for Regulating

;; TITLE OF INVENTION: Expression of Polypeptides and Its Use in the Preparation of  
;; TITLE OF INVENTION: 2,2-Dialkylglycine Decarboxylase of Pseudomonas Cepacia  
;; NUMBER OF SEQUENCES: 30  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
;; ADDRESSEE: Dunne  
;; STREET: 1300 I Street, N.W., Suite 700  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: US  
;; ZIP: 20005-3315  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/952.817  
;; FILING DATE: 19920928  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Meyers, Kenneth J.  
;; REGISTRATION NUMBER: 25,146  
;; REFERENCE/DOCKET NUMBER: 01120.0002-01000  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-408-4000  
;; TELEFAX: 202-408-4400  
;; INFORMATION FOR SEQ ID NO: 26:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 291 amino acids  
;; TYPE: AMINO ACID  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-07-952-817-26

Query Match 5.9%; Score 85.5; DB 1; Length 291;  
Best Local Similarity 21.8%; Pred. No. 0.18;  
Matches 63; Conservative 44; Mismatches 113; Indels 69; Gaps 11;  
QY 1 MNPQLDLSIIIEGSEFEGASLALSPSAVSORVKALEHHVGRVLVSRTPAKATEA 59  
Db 1 MNRIDLEYVALAHRFRRAADSCHVSQPTLSGQIRKLEDELGMILLERSRVLFQA 60  
QY 60 GEVLVQARKKVLQAETKAQLSGLAEI---PLTIAINADSLSTWFPVNEVASWGGA 116  
Db 61 GMLLVQAR-TLVREVKVLEKMSAQOGETMSGPLHGL---ITVGYLLPHIIPMLHQ 115  
QY 117 TLTLRLE---DEAHTLSLLRGG-----VLGAVTREA-----NP 147  
Db 116 TFP-KLEMYLHEAQTQLLAQDCKLDCVTLALVKESERFIEVPLFDEPMLAIYEDHP 174  
QY 148 VAGCEVVELGTMRLHAIATPSLRDAYVDGKLDWAAMPVLRFGPKDVLQDRDL----DGR 203  
Db 175 WANRECVMA-----DLAGEKLLMLEDHCHLRDQAMGCFEAG 212  
QY 204 VDGPGVRRRSIVPSABGFGBAIRGLGWGLLPETQAAPMLKAGEVILL 252  
Db 213 ADEDTHERATSL---ETLRNVAAGSGITLLPALAVPPERKRGVYL 257

RESULT 13  
US-08-537-210A-2  
; Sequence 2, Application US/08537210A  
; Patent No. 5780300  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Fortini, Mark  
; APPLICANT: Matsuno, Kenji  
; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY  
; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Pennie & Edmonds  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: NY  
;; COUNTRY: USA  
;; ZIP: 10036/2711  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/537,210A  
;; FILING DATE: 29-SEP-1995  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Mirock, S. Leslie  
;; REGISTRATION NUMBER: 18,872  
;; REFERENCE/DOCKET NUMBER: 7326-027  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 212-790-9090  
;; TELEFAX: 212-869-8864  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1068 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
;; FEATURE:  
;; NAME/KEY: Human N1 (TAN-1)  
;; LOCATION: 1152...2219  
;; OTHER INFORMATION: Highly conserved ankyrin repeat  
;; OTHER INFORMATION: region of No. 5780300ch  
;; US-08-537-210A-2

Query Match 5.8%; Score 85; DB 1; Length 1068;  
Best Local Similarity 21.6%; Pred. No. 1.5;  
Matches 70; Conservative 44; Mismatches 106; Indels 104; Gaps 16;  
QY 11 SIIDEGSEFEGASLALSPSAVSORVKALEHHVGRVLVSRTPAKATEAGEVLVQAARKM 70  
Db 758 AVISDFIYQASL-----HNQDRTGTALHLAAR--YSRSDAAKR----- 796  
QY 71 VLLQAEYKAQLSGLAEIPLTIAINADSLSTWFPVNEVASWGAT-LTLRLEDEAHTL 129  
Db 797 -LLEASADANTQDMNGRTPLHAAVSADAQGVF-----QILINRATDLDARMHDGTTPL 849  
QY 130 SLLRRGDVLGAVTREANPVAGCEVV-ELG-TMRHLAIATPSLRDAYVW----- 175  
Db 850 ILAARLAVEGMLEDLINSADVNAVDDLGKLSALHWAANNV-DAVVLLKNGANKDMQN 908  
QY 176 -----DGKLDWAAMPVLRFG-----PKDVLQDR---DLGGRVD----- 205  
Db 909 NREETPLFLAAWEGSYETAKVLLDFANWDITDHMDRLPRDIAQERMHHDIVRLLEYNL 968  
QY 206 -----GPGVRRRSIVP--SAEGF-----GEAIRRLGWGLLPETQAAPMLKA 246  
Db 969 VRSPQLHGAPLGGPTLSPPLCSNPGYLSLKGPGVGGKVRKPSKGLACGSKAKDLKA 1028  
QY 247 -----GEVILLDE---IPIDT 259  
Db 1029 WRKKSQDGKGLLDSSGMLSPVDS 1052

RESULT 14  
US-09-113-825-2  
; Sequence 2, Application US/09113825  
; Patent No. 6149902  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon





QY 247 -----GEVILLDE---IPIDT 259  
|: ||| |:  
Db 2180 WRKKSQDGKGLLDSSGMLSPVDS 2203

Search completed: September 13, 2001, 17:52:12  
Job time: 1118 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:52:41 ; Search time 25.14 Seconds  
(without alignments)  
878.704 Million cell update

```

Title: US-09-105-1171-3
Perfect score: 1460
Sequence: 1 MNPQLDTLLSIIDECSFEG.....RSLARLTDAVVDAIEGLRP 290

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```

```
Searched:      219241 seqs, 76174552 residues      .
Total number of hits satisfying chosen parameters: 219241
```

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```
Database :      PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	538.5	36.9	303	2	G70756	hypothetical prote
2	410	28.1	300	2	B83100	inhibitor of chro
3	376.5	25.8	297	2	S22098	replication initi
4	376.5	25.8	297	2	F85947	hypothetical prote
5	361.5	24.8	298	2	C82318	chromosome initi
6	177	12.1	297	2	G83145	probable transcri
7	175	12.0	300	2	D83502	probable transcri
8	174	11.9	295	2	E83302	probable transcri
9	170	11.6	310	2	E83381	probable transcri
10	163.5	11.2	286	2	A83482	probable transcri
11	156	10.7	304	2	E83495	probable transcri
12	152.5	10.4	292	2	G83332	probable transcri
13	152.5	10.4	295	2	D82443	transcription regu
14	151	10.3	302	2	B82036	transcription regu
15	149.5	10.2	339	2	C83290	probable transcri
16	148	10.2	303	2	G83354	probable transcri
17	148	10.1	309	2	S70535	probable transcri
18	143.5	9.8	297	2	D86063	positive regulator
19	143.5	9.8	305	2	F85823	nitrogen assimilat
20	142.5	9.8	305	2	D64963	nitrogen assimilat
21	141.5	9.7	311	2	H83174	probable transcri
22	140.5	9.6	305	2	B47099	nitrogen assimilat
23	140	9.6	312	2	E83624	probable transcri
24	139.5	9.6	297	1	RGECIV	regulatory protein
25	139	9.5	309	2	D64139	metR protein homol
26	138.5	9.5	305	2	G82167	transcription acti
27	138.5	9.5	306	2	T35595	probable transcri
28	138.5	9.5	308	2	H65014	hypothetical prote
29	138	9.5	293	2	H70067	transcription regu

## ALIGNMENTS

RESULT 1

G70756  
hypothetical protein Rv1985c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: G70756  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
J.; Connor, R.; Davies, K.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: G70756  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-303 <COL>  
A:Cross-references: GB:274025; GB:AL123456; NID:g3261586; PIDN:CAA98410.1; PID:g32615  
A:Experimental source: strain H37RV  
C:Genetics:  
C:Superfamily: conserved hypothetical protein H11364  
A:Gene: Rv1985c

Query Match	36.9%;	Score 538.5;	DB 2;	Length 303;
Best Local Similarity	43.1%;	Pred. NO. 7.8e-36;		
Matches 124:	Conservative	49;	Mismatches 106;	
				Indels 9;
				Gaps 5;

QY	5	QLODTLUSIIDBGSFEGASIALSTSPASQRYKALBHHVGRVULVSTQPAKATEAGEVLV	64
Db	10	QLAALAAVVGLSFGSFAAAERLHTTPSAVSQRKISLEQQQGVQLVREKPCRTATGIPLL	69
QY	65	QAARKMVLLOAETKAOLSGR-LAEIPLTIATNADSLSTWFPVFNEVASWGATLTLLRL	122
Db	70	RLAAQTALLSEBALAEEMGNASLKRTTITIAVNDSNATWFSNVPDGL--GDVLLDWRI	126
QY	123	EDEAHTLSLRRGDVLGAVTPEANPVAGCEVVELGPMRHLAIATPSLRDAYVMVDGKLDA	182
Db	127	EDQDHARLLREGVAMGAVTTERNPVPGCRVHPHLPGBMRYLPVASRPFVQRHLSDGFTAA	186
QY	183	A--MPVLRFCPKDVLDDRLDGDGVDCVPGRRRRSYVPSAEGGEATRRGLGKGLLPETCA	240
Db	187	AAKAPSLAWNRRDGLQDM-LVRKAFRAITRPTHVFVTEGTGTAARAGLGMGMFPEKLA	245
QY	241	APMLKAGEVILLDEIPIDTPMYQWRRLSRLARTDAVDDAAIEGL	288
Db	246	ASPLADGSFVRVCDIHLDPVLYWCKLSDPSPIARTID-TVRAAASGL	292

## RESULT 2

2  
B83100  
inhibitor of chromosome initiation IciA PA4363 {imported} - *Pseudomonas aeruginosa* (s

A;Accession: JN0079  
A;Molecule type: DNA  
A;Residues: 1-297 <THO>  
A;Cross-references: GB:M62865; PIDN:AAA62780.1; PID:g146436  
A;Accession: PS0203  
A;Molecule type: protein  
A;Residues: 1-28, 'X', 30-35, 'X', 37-38, 'XX', 41 <TH2>  
A;Blattnr: F.R.: Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of *Escherichia coli* K-12.  
A;Reference number: A64720; MUID:97426617  
A;Accession: C65076  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-297 <BIAT>  
A;Cross-references: GB:AE000375; GB:U00096; NID:g1789282; PIDN:AACT5953.1; PID:g17892  
A;Experimental source: strain K-12, substrain MG1655

C;Genetics:  
A;Gene: icia  
A;Map position: 62.8 min  
C;Function:  
A;Description: binds to the origin (oric) of Escherichia coli DNA to block initiation  
C;Superfamily: regulatory protein ampr  
C;Keywords: DNA binding; DNA replication inhibitor  
F;21-40/Region: helix-turn-helix motif

Query Match            25.8%; Score 376.5; DB 2; Length 297;  
Best Local Similarity 34.1%; Pred. No. 7.3e-23;  
Matches 100; Conservative 53; Mismatches 125; Indels 15; Gaps 3;

QY     6 LDYLLSIIDSGSPGASLALISPSAVSQRVKALEHHVGRVLRSRTPAKATAGEVLVQ 65  
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db     9 LQALDAVIREGFERRFAQAQLKITQSASQRIKOLENNFGQPLLVRVPVPRPTGGQKLLA 68  
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY     66 AARKWILLQAEATKAQLSRLAEPLTTAINADSLSTWFPPVFVEVASWGCGATLTLRLEDE 125  
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db     69 LLRKVELLEEENLGDEGTGSTPDLLSLVAVNADSLWTLLPALAPVLADSIRINLQVEDE 128  
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY     126 AHYLLSLRRRGDVLGCAVTREANPVAGCEVVELGTMRHLAIATPSLRDAYMYMDGKLDAAM- 184  
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db     129 TRTQERLRRCGVVGAVSIGIQHQLPSCLVDKLGALDYLFVSSKKPFAEKYFPNGVTRSAALK 188  
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```

185 -PVLRFQPKDVLQDRDLGRVDGPGVRRRRYSIVPSAEGFGEAIRRGLGWGLLPETAAPM 243
  ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 189 APVAFDHLDMHQAIFLQNFEDLPPGSVPCHVNSSFAFVLQARQGTTCMIPHQLQIEK 248

QY 244 LKAGEVLLDEIPIDTP-----MTWQRWRESRLARLTDAVDAATGLR 289
  | : || | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 249 LASGELIDL-----TPGLFQRMLYWHRPAPESRMKRVTDALLDYGHKVL 295

RESULT 4
F85947
hypothetical protein icia [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: F85947
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grobeck, E.J.; Davis, N.W.; Llin, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85947
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <STO>
A:Cross-references: GB:AE005174; NID:g12517445; PIDN:AAG58042.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: icia

```

RESULT 5  
 C82318  
 chromosome initiation inhibitor VC0482 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)  
 C:Species: *Vibrio cholerae*  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C:Accession: C82318  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 Chardonson, D.; Ermolaeva, M.B.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.  
 L., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
 A:Reference number: A82035; MUID:20406833  
 A:Accession: C82318  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-298 <HEI>  
 A:Cross-references: GB:AE004134; GB:AE003852; MID:g9554900; PIDN:AAF93655.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VC0482  
 A:Map position: 1

RESULT 7  
D83502 probable transcription regulator PA1141 [imported] - Pseudomonas aeruginosa (strain P  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: D83502  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MUID:20437337  
A:Accession: D83502  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-300 <STO>  
A:Cross-references: GB:AE004544; GB:AE004091; NID:g9947060; PIDN:AAG04530.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA1141  
C:Superfamily: hypothetical protein b1875

Query Match	Score	DB 2:	Length
	12.0%	175:	300:



A:Gene: PA1309

Query Match 11.2%; Score 163.5; DB 2; Length 286;  
Best Local Similarity 27.7%; Pred. No. 8.1e-06;  
Matches 74; Conservative 32; Mismatches 126; Indels 35

Qy	1	MNPQTDLTSLIDEGSFEGLASLSPGAVSQRVKALEHHVGRVLVSRT-QPAKATEA	59
Dd	1	MNLFQRAFDVAAREGSFTAAERLCISQPAVTGHVKALAEHYQVTLRFETARRILEE	60
Qy	60	GEVLVQAARKMVLQAKTKAQLSGLAEIPLTTAINADSLSTWFPVPVENEVASWGATLT	119
Dd	61	GSRLVAITRTLFALEEEAEALLDANROLVSGRLVEAGPHLVMPMLARLRASYPGITVN	120
Qy	120	LRLDEAHITLSLLRRGDVLGAVTREAMPVAGEVVELGTMRHLIAITAIPSDRAYMYVDGKL	179
Dd	121	LRLNGAETHLAALSSEHVDVAVLITEVARPGLGLEELVNSRICALLPRQHPWCERNEGV-	179
Qy	180	DWAAMPVLRFGPKDVLDQRDLDDGVDPGVGRR-----RVSI-VPSAEGFGEA	225
Dd	180	-----PLEELDQQIMVLREPESITRRTFDRCABEQGVQRPVLLDLSREAVTEA	228
Qy	226	IRRLGLGWGLL-----PETQAAPML	244
Dd	229	VAAELGGVYSSLLEVGNDRPVRAVPLV	255

RESULT 11

E83495  
 Probable transcription regulator PA1201 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
 C:Species: *Pseudomonas aeruginosa*  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: E83495  
 R:Stover, C.K.; Pham, X.-Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, K.R.; Kas, A.; Larbig, K.; Lim,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho  
 A:Reference number: AB2950; MUID:20437337  
 A:Accession: E83495  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-304 <SFO>  
 A:Cross-references: GB:AE004550; GB:AE004091; NID:g9947122; PIDN:AAG04590.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA1201  
 C:Superfamily: conserved hypothetical protein H11364

Query Match 10.7%; Score 156; DB 2; Length 304;  
Best Local Similarity 26.7%; Pred. No. 3.5e-05;  
Matches 76; Conservative 52; Mismatches 109; Indels

Qy	6	LDTLLSIDSGSPGASLALISPSAVSQVRKALEHHVG--RVLVSSTQPAKATEAGEVILV	64
Db	8	IQAFLNVVLQSSIAAERNLSKSVISKRVSOLERHGLVRLLYSTRVNETPTEAGGFFY	67
Qy	65	QAARKMYLLQ-----AETKA-----QLSGRLAET-PUTIAINADSLSTWPPPVENE--VASW	113
Db	68	KSAK--ASLQDLNNAAESVALRENDLCGELRVMAPNSFG-----TLWLGLPLVMEPMARN	119
Qy	114	GGATLTLRLEDEAHTLSLLRRG--DVLGAVNR--EANPVAAGVEVLGTRHRLAIATPSLR	170
Db	120	PRLEVLQLQD--RIVDFEKEGFDLAIRITRLQDSSLIAR-----QLGTSRRVVCCSPEYL	173
Qy	171	DAYMVDGKL---DWAAMPVLRFGPKDVLQDRLDGRVDG-----PYGRRRVSTVPSA	219
Db	174	ERH---GPLQRIEDILCHPICGYSHNTPSOLWSFEPRVAGAPARMITPRGRFNTN---NG	227
Qy	220	EGFGEAATRRGLGGLLPETCAAPMLKAGEVI---LLDEIPIDTPMY	262

Db 228 QTMRDAAVRGLGLAMLPLFIAAEDLAAGRLVEALPQERPLDDLIY 272

## RESULT 12

G83332  
Probable transcription regulator PA2497 [imported] - *Pseudomonas aeruginosa* (strain P  
C:Species: *Pseudomonas aeruginosa*  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83332  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa  
A:Reference number: A82950; MUID:20437337  
A:Accession: G83332  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-292 <SFTO>  
A:Cross-references: GB:AE004677; GB:AE004091; NID:g9948548; PIDN:AAG05885.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2497  
C:Superfamily: hypothetical protein b1875

Query Match 10.4%; Score 152.5; DB 2; Length 292;  
Best Local Similarity 25.9%; Pred. NO. 6.4e-05;  
Matches 84; Conservative 41; Mismatches 122; Indels 77

Qy	4	IQLDTLSIIIDEGSFEGLASLSTSPSAVSORVKALEHHGVILVSRQ--PAKATEAGEV	62
Db	5	VLLTFVSVDTGNTRAGEHLHLTQSTFVSQQVIRLEQNLGCRLLDRSQVLPTEEGR	64
Qy	63	LVQARKNVLLQAEKQALSGRLAEIPLTAINADSLSTWFPFVNEVASGGGATLTRL	122
Db	65	LLGYARLLRLUSDASBALSPAHGEGGVLRIGVPEDLAGEVLMPLVTRFTE---ERPRLRL	121
Qy	123	EDEA-----HTLSLLRRGDVLGAVTREANPVAGCSEVVELGTMRHLLAIATPSLRDAYMVDGK	178
Db	122	EVEGSLSHLLRLYRSGELDLLLWKQAGSDC-----HARWAEP-----	161
Qy	179	LDW---AAMPVLRGPKDVLQDRDLGDVDPVGR--RVSIVPSAEFGFGE-----	224
Db	162	LGNFESAARPGEGSGPEPVP-----LVVFPVGAALYRQEMIIHALESIGRRWRIGVSSAS	215
Qy	225	-----AIRRGLGWLGLP-----ETQAA--PMLKAGEVILLDEIPIDTPMYQWR	266
Db	216	LASIVAAVGAGLGVSLPLGICVCGPEHRLIGAAOAGFPPTAGIELALYARPELDS-----	268

RESULT 13

D82443  
transcription regulator LysR family VCA0575 [imported] - *Vibrio cholerae* (strain N169)  
C:Species: *Vibrio cholerae*  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: D82443  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.  
Charadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers  
L, R.R.; McKelanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A:Reference number: A82035; MUID:20406833

A;Accession: D82443

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-295 <HEI>  
A:Cross-references: GB:AE004388; GB:AE003853; NID:g9657979; PIDN:AAF96477.1; GSPDB:GN  
A:Experimental source: serogroup O1: strain N16961: biotype El Tor

## C:Genetics:

A:Gene: VCA0575

A:Map position: 2

C:Superfamily: probable transcription regulator ybbs

Query Match 10.4%; Score 152.5; DB 2; Length 295;  
Best Local Similarity 25.5%; Pred. No. 6.4e-05;  
Matches 79; Conservative 62; Mismatches 102; Indels 67; Gaps 17;

QY 1 MNPIDLTLLSIID----EGSFGASLALSISPSAVSQRVKALEHHVGRVLSRT--QPAK 55  
DB 2 LSPITILEA-LHLIDATERGSGFAAANEINRAPSSLSYQIQLEQDLIMIFDRSHRAN 60  
QY 56 AYEAGEVLVQAARKMVLQAEYK---AOLSGRLAIPITIAINADSLSTWFFPP---VEN 108  
DB 61 FEAGKLIILGR--AILATEKLVNDATYLLANGWELDTITIALDG-----IVPAANLFP 112  
QY 109 EVASWGATLT-LRLEDE--AHTLSLLR--RGDVLGAVTREPANPVAGCEVVELGTMRLHA 163  
DB 113 MVEALGNISKTRVRIQDEILAGWEALATGRADLLICPRIEALP-QDVKAETIGTKMIW 171  
QY 164 IATPS-----LRDAY----WYDGKLDWAAMPVLRFGPKDVLQDRDLGRVDG 206  
DB 172 VAAPIHYVHRRSGEFNEAREKRYAIAIADTAREQPMASV-----NILQ----- 215  
QY 207 PVGRRVSVPSAEGFGEAIRRGLGMLLPETQAAAPMLKAGEVILL---DEIPIDTPMYW 263  
DB 216 ---RQPLTVSNLDKAKKALVAGLGIGTLPLOVAQPYIDKGLKAIHGSDELMEDIVLAN 272  
QY 264 QRWRL-ESRS 272  
DB 273 RRNQGEAKS 282

## RESULT 14

B82036

Transcription regulator lysr family VC2760 [imported] - Vibrio cholerae (strain N16961)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: B82036

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: B82036

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-302 &lt;HEI&gt;

A:Cross-references: GB:AE004342; GB:AE003852; NID:g957358; PIDN:AAF95899.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC2760

A:Map position: 1

C:Superfamily: probable transcription regulator ybbs

Query Match 10.3%; Score 151; DB 2; Length 302;  
Best Local Similarity 26.1%; Pred. No. 8.8e-05;  
Matches 69; Conservative 53; Mismatches 118; Indels 24; Gaps 10;

QY 6 LDTLLSIIDEGSFGASLALSISPSAVSQRVKALEHHVGRVLSR--TOPAKATEAGEVLV 64  
DB 8 LEMLDVARLSGFTAAAEVLHVLPASISYVRQISELGVLFRLPRKVELTPAGELFM 67  
QY 65 ----QAARKMVLQAEYK---AOLSGRLAIPITIAINADSLSTWFFPPVFNVEASWGATLT 120  
DB 68 LEAROLLRQMEIRIAQTKRAAGWKRKTLKVTLD--NVVKDKMKPMVEAFYQTFDFAEIQI 126  
QY 121 RLEDEAHTLSLLRGD---VLGAVTREPANPVAG-CEVVELGTMRLHAATATPS---LRDAY 173

DB 127 NMEVFNGSWEAIAQGRADIVIGATA--AVPVGSGDFEVRDMGILDWAFVMSPNHPCVREQN 184  
QY 174 MYDGKLDWAAMPVLRFGPKDVLQDRDLGRVDGPGVRRRSIVPSAEGFGEAIRRGLGWS 233  
DB 185-LSEGFIIS-OFLAICLDTSSTVLPKRHTE-----HYPKQRLLLPWNWYSAIESLSKSLGIVG 238  
QY 234 LLPETOAAAPMLKAG---EVILLDE 254  
DB 239 YMPRHMAQPLLASGQLVEXVLPDE 262

## RESULT 15

C83290

probable transcription regulator PA2848 [imported] - Pseudomonas aeruginosa (strain P

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: C83290

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: C83290

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-339 &lt;STO&gt;

A:Cross-references: GB:AE004711; GB:AE004091; NID:g9948927; PIDN:AAG06236.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2848

C:Superfamily: conserved hypothetical protein H11364

Query Match 10.2%; Score 149.5; DB 2; Length 339;  
Best Local Similarity 27.2%; Pred. No. 0.00013;  
Matches 83; Conservative 40; Mismatches 131; Indels 51; Gaps 13;

QY 7 DPL-----LSIIDEGSFGASLALSISPSAVSQRVKALEHHVGRVLSRTQPA-KATE 58  
DB 25 DTLSEQFVLYLDVLDGGSFSAARKHPLTPSAVARRMDALERAVGSTLLVTRTHAVRATP 84  
QY 59 ACEVLVQAARKW----LLQATKALQSG--RLAEIPLTIANADSLSTWFFPPVNEVAS 112  
DB 85 AGLAFADRARRIVTELRLARAEAVSLSTAPOGLIRIDAPVFGRRHLA---PAVADFLKA 141  
QY 113 WGGATLTLRLEDEAHTLSLLRGDVLGAV--TREANPVAGCEVW--ELGTMRLHAATATPS 168  
DB 142 NPGLDVQLRLIGSFIDL-----QGEHLGEVDLVLRAGLPUDSRSLVATSLAPMVVVVCASPE 197  
QY 169 LRDAYMVDGK-----LDWAAMP---VLRFGPKDVLQDRDLGRVDGPGVRRRSIV 216  
DB 198 YLREHGVASPGELPEHAGIDWNLSPYAWRF-----QHDGKLOHLRPRKARLAT 248  
QY 217 PSAEGFGEAIRRGLGMLLPETQAAAPMLKAGEVILL---DEIPIDTP---MYQWRRLSES 271  
DB 249 NNAEAMWDAALAGLIGIAHLPTWLCSEYLLRGELQALFCDGDLPAAEPTCIYALRLERAS 308  
QY 272 SLARL 276  
DB 309 SRTSL 313

Search completed: September 13, 2001, 17:52:42

Job time: 1067 sec



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.  
OM protein - protein search, using sw model  
Run on: September 13, 2001, 17:57:09 ; Search time 15.26 Seconds  
(without alignments)  
650.989 Million cell updates/sec  
Title: US-09-105-117i-3  
Perfect score: 1460  
Sequence: 1 MNPQLDITLLSIIDEGSFEG.....RSLRLTDAVVDAIEGLRP 290  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 93435 seqs, 34255486 residues  
Total number of hits satisfying chosen parameters: 93435  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1460	100.0	290	1	LYSG_CORGL
2	538.5	36.9	303	1	YJ85_MYCTU
3	376.5	25.8	297	1	ICIA_ECOLI
4	365.5	25.0	297	1	ICIA_EDWIC
5	355.5	24.3	299	1	ICIA_AERSA
6	148	10.1	309	1	LTRA_KLEPN
7	142.5	9.8	305	1	NAC_ECOLI
8	140.5	9.6	305	1	NAC_KLEAE
9	139.5	9.6	297	1	ILVY_ECOLI
10	139	9.5	309	1	METR_HAEIN
11	138.5	9.5	308	1	YFER_ECOLI
12	136.5	9.3	317	1	METR_SALTY
13	136	9.3	294	1	LRRA_SYN7
14	129.5	8.9	317	1	METR_ECOLI
15	128	8.8	295	1	TEPR_ALCEU
16	127.5	8.7	301	1	ESTR_ACTICA
17	127.5	8.7	311	1	OXTR_MYCLE
18	127.5	8.7	312	1	LRHA_ECOLI
19	127.5	8.7	326	1	BLAA_STRCI
20	126.5	8.7	309	1	NTCB_SYN7
21	126.5	8.7	317	1	CFXR_ALCEU
22	126	8.6	316	1	PECT_ERWCH
23	125.5	8.6	317	1	YC30_PORPU
24	121	8.3	310	1	YDHB_ECOLI
25	117.5	8.0	298	1	TRPI_PSESY
26	116.5	8.0	306	1	OPRR_PSEAE
27	115.5	7.9	298	1	CBRR_RHOU
28	115.5	7.9	311	1	OXTR_MYCAV
29	111.5	7.6	307	1	YEAT_ECOLI
30	111	7.6	292	1	HVRB_RHOCA
31	111	7.6	299	1	YCJZ_ECOLI
32	108.5	7.4	308	1	YBBS_ECOLI
33	108	7.4	312	1	YMR2_MYCTU

34	107.5	7.4	298	1	YHAI_ECOLI	P42623	escherichia
35	107.5	7.4	309	1	NTCB_SYN3	P74422	synecocyst
36	107	7.3	292	1	ILVY_HAEIN	P44821	haemophilus
37	107	7.3	298	1	Y4WQ_RHISN	P55576	rhizobium s
38	107	7.3	304	1	TU44_AGRVI	P52669	agrobacteri
39	106.5	7.3	293	1	YKUM_BACSU	O34827	bacillus su
40	106.5	7.3	294	1	CLCR_PSEPU	Q05840	pseudomonas
41	106.5	7.3	299	1	YWFK_BACSU	P39847	bacillus su
42	106.5	7.3	308	1	YTLI_BACSU	O35038	bacillus su
43	106.5	7.3	310	1	CBRR_RHOSH	P52690	rhodobacter
44	105.5	7.2	143	1	BUDR_ENTAE	P52665	enterobacte
45	104.5	7.2	299	1	CYNR_ECOLI	P27111	escherichia

ALIGNMENTS

RESULT 1

LYSG\_CORGL STANDARD; PRT; 290 AA.

AC P94632;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE LYSINE EXPORT TRANSCRIPTIONAL REGULATORY PROTEIN LYSG.

GN LYSG.

OS Corynebacterium glutamicum (Brevibacterium flavum).

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;

OC Corynebacterium.

OX NCBI\_TaxID=1718;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=R127;

RX	MEDLINE=97126810;	PubMed=8971704;
RA	Vrljic M., Sahm H., Eggeling L.;	
RT	"A new type of transporter with a new type of cellular function: L-	
RT	lysine export from Corynebacterium glutamicum."	
RL	Mol. Microbiol. 22:815-826(1996).	
CC	-!- FUNCTION: PROBABLE TRANSCRIPTIONAL REGULATOR OF LYSG.	
CC	-!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL	
CC	REGULATORS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement. (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
CC	EMBL; X96471; CAA65323.1; -	
DR	InterPro; IPR000847; -	
DR	Pfam; PF00126; HTH\_1; 1.	
DR	PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.	
KW	DNA-binding; transcription regulation.	
FT	DNA\_BIND 18 37 H-T-H MOTIF (POTENTIAL).	
SQ	SEQUENCE 290 AA; 31388 MW; 9CE0A63F775FCB74 CRC64;	

Query Match 100.0%; Score 1460; DB 1; Length 290;

Best Local Similarity 100.0%; Pred. No. 6.4e-110;

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNPQLDITLLSIIDEGSFEGASLALSISSAVSQRVKALEHHVGRVLVSRTPAKATEAG 60

|||||

Db 1 MNPQLDITLLSIIDEGSFEGASLALSISSAVSQRVKALEHHVGRVLVSRTPAKATEAG 60

|||||

QY 61 EVLVQAARKMVLQAETKAQLSGRLAEIPLTIAINADSLSTWPPVFNVEASVGGATLTL 120

|||||

Db 61 EVLVQAARKMVLQAETKAQLSGRLAEIPLTIAINADSLSTWPPVFNVEASVGGATLTL 120

|||||

QY 121 RLEDEAHTLSLLRRGDVLGAVTREPANPVACEVVEGLTMRHLAIPSLRDAVWDGKLD 180

123	QY	EDAH	TL	SL	LR	RG	DV	LG	AV	TR	EA	NP	AV	GE	VE	LG	TM	RH	LA	TA	TS	LR	DA	YV	WD	GK	LD	WA	182
127	Db	EDDH	SAR	LL	REG	VAM	GA	VT	TE	RN	PG	CR	VP	LG	EM	RY	LP	VA	SR	PF	VQ	RH	LS	DG	FT	AA	186		
183	QY	A	-	MP	VL	RG	PD	VL	DD	RG	RV	DG	PG	VR	RR	VS	IV	PS	AE	GE	AT	RG	LG	WG	LL	PE	TA	240	
187	Db	AK	AP	SL	AW	NR	DD	GL	QM	-	LV	RA	FR	AI	TR	PH	TF	VT	TE	GT	TA	AR	AG	LG	MG	MP	PE	KL	245
241	QY	AP	ML	KAGE	V	ILL	DE	TP	ID	TP	MY	QW	RW	LR	ES	SL	AR	UT	DA	VD	AA	1E	GL	288					
246	Db	ASP	LAD	GS	FVR	VC	DI	HL	DY	PL	YQ	WC	WK	LD	SP	II	AR	TD	-	TV	RA	AS	GL	292					

RESULT	3				
ID	ICIA_ECOLI	STANDARD;	PRT;	297 AA.	
AC	P24194;				
DT	01-MAR-1992	(Rel. 21, Created)			
DT	01-MAR-1992	(Rel. 21, Last sequence update)			
DT	01-NOV-1997	(Rel. 35, Last annotation update)			
DE	CHROMOSOME INITIATION INHIBITOR (ORIC REPLICATION INHIBITOR).				
DE	CHROMOSOME INITIATION INHIBITOR (ORIC REPLICATION INHIBITOR).				
GN	ICIA.				
OS	Escherichia coli.				
OS	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Escherichia.				
OC	Escherichia.				
OX	NCBI_TaxID=562;				
RN	[1]				
RC	SEQUENCE FROM N.A., AND SEQUENCE OF 1-41.				
RC	STRAIN=K12 / W3110;				
RC	MEDLINE=91239508; PubMed=2034653;				
RX	Thoeny B., Hwang D.S., Fradkin L., Kornberg A.;				
RT	"Icia, an Escherichia coli gene encoding a specific inhibitor of				
RT	chromosomal initiation of replication in vitro."				
RL	Proc. Natl. Acad. Sci. U.S.A. 88:4066-4070(1991).				
RP	[2]				
RP	SEQUENCE FROM N.A.				
RP	STRAIN=K12;				
RC	Roy I., Leadlay P.F.;				
RA	Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.				

RC STRAIN-K12 / MG1655;  
RX MEDLINE-97426617; PubMed-9278503;  
RY Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Gosden M.A., Rose D.J.,  
RY Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12.";  
RT Science 277:1453-1474(1997).  
RN [4]

```

RN      [4]
RP      SEQUENCE OF 1-176 FROM N.A.
RC      STRAIN-K12;
RX      MEDLINE=93374859; PubMed=8366047;
RA      Hove-Jensen B., Maigaard M.;
RT      "Escherichia coli rpiA gene encoding
RL      J. Bacteriol. 173:5628-5635(1993).
CC      - !- FUNCTION: SPECIFIC INHIBITOR OF
CC      REPLICATION IN VITRO. BINDS THE THREE 13-MERS IN THE ORIGIN
CC      (ORIC) TO BLOCK INITIATION OF REPLICATION.
CC      - !- SUBUNIT: BEHAVES AS HOMODIMER IN SOLUTION.
CC      - !- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC      REGULATORS.

```

-!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M62865; AAA62780.1; .

```

DR EMBL; X66836; CAA47310.1; -.
DR EMBL; U28377; AAA69083.1; -.
DR EMBL; AE000375; AAC75953.1; -.
DR EMBL; X73026; CAA51508.1; -.
DR PIR; JN0079; JN0079.
DR PIR; S22098; S22098.
DR Ecogene; EG10490; icia.
DR InterPro; IPR000847; -.
DR Pfam; PF00126; HTH_1; 1.
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW DNA-binding; Transcription regulation.
FT DNA_BIND 21 40 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 297 AA; 33471 MW; 8AD60B7B66G61E3EF CRC64;

Query Match          25.8%; Score 376.5; DB 1; Length 297;
Best Local Similarity 34.1%; Pred. No. 3.5e-23;
Matches 100; Conservative 53; Mismatches 125; Indels 15; Gaps 3;

QY      6 LDTLISIDEGSFEGASIALSISPVSAYSQRVKALEHHVGRVLVSGRTOPAKATEAGEVLVQ 65
       + : : | | | | : : : : | | | | | : : : : | | : : | | : :
Db      9 LQALDAVIRGERGFERRAQAQKLCITOSAVSQRIKOLEMFGQPQLLVRTVPPTPEQGKKLLA 68

QY     66 AARKWVLQAEKTKAQLSGRLAEIPLTAINADSISTWFPPVFNEVASWGGAATLILRLEDE 125
       + : : | | : : : : : : | | | | | : : : : | | : : | | | |
Db     69 LLRGVELLEEWEGLDEQTGSTPLLSLAVNADSLATWLPLPALAPVLADSPTRLNLQVEDE 128

QY    126 AHTLSLRGGDVLGAVTREANPVAGCEVVELGTWRHLAIATPSLRDAYMVDGKLDWAAM- 184
       + | | | | | : | | | | : : : : | | : : | | : : | | : :
Db   129 TRTOELRRGEVCAVSIHQALPCLVDKLGALDYLFVSSKPAEKYFPNGVTFRSALLK 188

QY   185 -PVLRFPGPKVDLQDRDLGRVDGPVGRRRSIVPSAEGFGEAIRGLGWGLLPETQAPM 243
       || : : | | | | : : | | | | | : : | | : : | | : : | |
Db  189 APVAFDHLDMDHAFLQONEDLPFGSPVCCHIVNSSEAFVQLARQGTCCMPHLQIEKE 248

QY   244 LKAGEVILLDEIPTDP-----MYQWRWLESRSRLRLTDVAVDDAEIGLR 289
       | : | | | | | | | | : : | | : : | | : : | | : : | |
Db  249 LASGELIDL-----TPGLFORMLYHRFAPEFSRMKRKTALLDYGHKVLR 295

RESULT 4
ICIA_EDWIC
ID ICIA_EDWIC STANDARD; PRT; 297 AA.
AC O52399;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CHROMOSOME INITIATION INHIBITOR (ORIC REPLICATION INHIBITOR).
GN ICIA.
OS Edwardsiella ictaluri.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Edwardsiella.
OX NCBI_TaxID=67780;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=93-146;
RL Moore M.M., Fernandez D.H., Thune R.L.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SPECIFIC INHIBITOR OF CHROMOSOMAL INITIATION OF
CC REPLICATION IN VITRO. BINDS THE THREE 13-MERS IN THE ORIGIN
CC (ORIC) TO BLOCK INITIATION OF REPLICATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to licensee@isb-sib.ch).

```

```

DR EMBL; AF037440; AAB92569.1; -.
DR InterPro; IPR000847; -.
DR Pfam; PF00126; HTH_1; 1.
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS00039; HTHLYSR_FAMILY; 1.
KW DNA-binding; Transcription regulation.
FT DNA_BIND 21 40 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 297 AA; 33392 MW; 93BEC456B2419989 CRC64;

Query Match 25.0%; Score 365.5; DB:1; Length 297;
Best Local Similarity 33.4%; Pred. No. 2.6e-22;
Matches 98; Conservative 51; Mismatches 129; Indels 15; Gaps 3;

Qy 6 LDTLSIIDESFEGASLALSISSAVSQRYKALEHHVGRVLVSRTPAKATEAGEVLVQ 65
   1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : :
Db 9 LQALDAVIRGERGFERAAKLCITQSAVSORIKOLENLFQGPLLVRTVPPRTEQGORLLA 68

Qy 66 AARKMVLQAEKTAQLSGRLAIEPIPTIAINADSLSTWFPVPFNVASGGATLRLLEDE 125
   : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : :
Db 69 LLHQVELLEEEWGLNDNSDGPLLLSLAVNADSLATLWLLPALKPYLADSPURLNLQVEDE 128

Qy 126 AHTLSLRGDLVGLAVTREANPVACEVVELGTMRLHAIATPSLRDAYMVDG--KLDWAA 183
   1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : :
Db 129 SRTQERLARGEVGAVSLQPLPCLVDRLGALDYLFCASPAFAARYFPNGVTSSLLK 188

Qy 184 MPVLRFQPKDVLQDRDLGVRGVPGRRRVSTVPSAEGFGAIRRGLGWGLLPETQAAAPM 243
   : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : :
Db 189 APAVAFDHLDMHQAFLLQNFELPGSPVFCVINSSEAFVQLAQLTCCMLPHLQVERE 248

Qy 244 LKAGEVILLDEPIDTP-----MYQWRWLESRLARLTDAVYDAIEGLR 289
   1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : :
Db 249 LRKGVLLIDL-----TPGVWRMMLYHRRFAPESLMMRRVTDALISHGRDVL 295

RESULT 5
ICIA_AERSA
ID ICIA_AERSA STANDARD; PRT; 299 AA.
AC P70773;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CHROMOSOME INITIATION INHIBITOR (ORIC REPLICATION INHIBITOR).
GN ICIA.
OS Aeromonas salmonicida.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxID=645;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 1102;
RA Swift S., Karlyshev A.V., Fish L., Durant E.L., Winson M.K.,
RA Williams P., Macintyre S., Stewart G.S.A.B.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SPECIFIC INHIBITOR OF CHROMOSOMAL INITIATION OF
CC REPLICATION IN VITRO. BINDS THE THREE 13-MERS IN THE ORIGIN
CC (ORIC) TO BLOCK INITIATION OF REPLICATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U65741; AAB70016.1; -.
DR InterPro; IPR000847; -.
DR Pfam; PF00126; HTH_1; 1.
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS00034; HTH LYSR_FAMILY; 1.

```

KW DNA-binding; Transcription regulation.  
 FT DNA\_BIND 21 40 H-T-H MOTIF (POTENTIAL).  
 SQ SEQUENCE 299 AA; 33410 MW; 5EEBECA44CC5A68D CRC64;

Query Match 24.3%; Score 355.5; DB 1; Length 299;  
 Best Local Similarity 33.6%; Pred. No. 1.7e-21;  
 Matches 98; Conservative 49; Mismatches 126; Indels 19; Gaps 6;

QY 11 SIIDEGSFEGASLALSISPSAVSQRYKALEHHVGRVLSRTPAKATEAGEVLVQAARKM 70  
 DB 14 AVMOQFNERAARLHTQSAISRIKLEQOFAEPLLRISQPLQATPLGOKLLAHYQV 73  
 QY 71 VLLQAEKALSGRL-----AETPLTAINADSLSTWFPVNEVASWGGATLTLED 124  
 DB 74 RQLELE-----LAGETAPDEPOAPTRVSTAVNADSLATWFLPALAPLEHQHPIELNLLVDD 129  
 QY 125 EAHTLSLLRRGDLGAVTREPANVAGCEVVELGTMRHLAIATPSLRDAYMVDGKLDNAAM 184  
 DB 130 ECRITLDRVREGQAFGAVSLHGOPLAGCCVDELGMRYLLTASPAFVARHFPAG-LTPAAL 188  
 QY 185 ---PVLREGPKDVLQDRDLGRVDGPGVRRRVSIVPSAEGEAGEAIRGLGGLLPETQAA 241  
 DB 189 AKTFAVAFDQDRDHVMFPMARHFGLEPGGYPCHTVRSSEAFVMAEQGLAYCLIPQLQIR 248  
 QY 242 PMLKAGEVILLDEIP---IDTPMYWQRWRLESRLARLTDVAVDAAIETGLRP 290  
 DB 249 QQLAOG--ILLDLSPSHLIEFLYHWRVLERGLHKLQISQRLISEGRALQP 298

RESULT 6  
 LTRA\_KLEPN  
 ID LTRA\_KLEPN STANDARD; PRT; 309 AA.  
 AC P52689;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE PROBABLE TRANSCRIPTIONAL REGULATOR LTRA.  
 GN LTRA.  
 OS Klebsiella pneumoniae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Klebsiella.  
 OX NCBI\_TaxID=573;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13882;  
 RX MEDLINE=96342382; PubMed=8748036;  
 RA Bott M., Meyer M., Dimroth P.;  
 RT "Regulation of anaerobic citrate metabolism in Klebsiella pneumoniae";  
 RL Mol. Microbiol. 18:533-546(1995).  
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; U31464; AAC44735.1;  
 CC DR InterPro; IPR000847;  
 CC DR Pfam; PF00126; HTH\_1.1;  
 CC DR PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.  
 KW Transcription regulation; DNA-binding.  
 FT DNA\_BIND 21 40 H-T-H MOTIF (POTENTIAL).  
 SQ SEQUENCE 309 AA; 34357 MW; 5C54D416605C734A CRC64;

Query Match 10.1%; Score 148; DB 1; Length 309;  
 Best Local Similarity 26.3%; Pred. No. 7e-05;

Matches 80; Conservative 50; Mismatches 126; Indels 48; Gaps 15;

QY 6 LDTLLSIIDEGSFEGASLALSISPSAVSQRYKALEHHVGRVLSR-TQPAKATEAGEVLV 64  
 DB 9 LALFQIVDQGSFSAVARQNGITPSAVSRVSRLEREMGCKLLQRTTRKRLSDAGETIY 68  
 QY 65 QAARKMV--LQAEKALSGRLAEIPLTIAI-NADSLSTWFPVNEVASWGGATLTLR 121  
 DB 69 QAOQMLEAARQAMDSAGSRQTVAGGKLTLSVPKAVGFRVHPLMMAFFHRYPOVDVCLR 128  
 QY 122 LEDAHTLSLLRRG-DVLGAVTREPANVAGCEVVELGTMRHLAIATPSLRDAYMVDGKLD 180  
 DB 129 LED--RPLDFDDGIDLALRITDTPSP--GLHGKPLMPIRHVICAT----EAY-----LQ 175  
 QY 181 WAAMPVLRFGPKDV-----LQDRDLGRVDGPGVRRRVSIVPSAEGFG----- 223  
 DB 176 QHGTP---YTQDLRAHSCISLGETPADARK----FRREGKTEVTQTYRYAANHATVR 228  
 QY 224 -EATRRGLGWLGPETQAAPMLKAGEVI-LLDE---IPIDTPMYWQRWRLESRLARLTD 278  
 DB 229 LDVAVRQHLGIGSLPLFTAREALANGDIQVLPWEFFISSYSDLLWAGDKHMPARMR- 287

RESULT 7  
 NAC\_ECOLI  
 ID NAC\_ECOLI STANDARD; PRT; 305 AA.  
 AC Q47005;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE NITROGEN ASSIMILATION REGULATORY PROTEIN NAC (NITROGEN ASSIMILATION CONTROL PROTEIN).  
 GN NAC.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=98155143; PubMed=9495755;  
 RA Muse W.B., Bender R.A.;  
 RT "The nac (nitrogen assimilation control) gene from Escherichia coli.";  
 RL J. Bacteriol. 180:1166-1173(1998).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 CC [3]  
 CC SEQUENCE FROM N.A.  
 RP STRAIN=K12;  
 RC MEDLINE=97251358; PubMed=9097040;  
 RX Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,  
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,  
 RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,  
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;  
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
 RL corresponding to the 40.1-50.0 min region on the linkage map";  
 CC DNA Res. 3:379-392(1996).  
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR FOR THE HUT, PUT AND URE  
 CC OPERONS AND REPRESSOR FOR THE GDH AND GLTB OPERONS IN RESPONSE TO

CC NITROGEN LIMITATION. NEGATIVE REGULATOR OF ITS OWN EXPRESSION (BY  
CC SIMILARITY). BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U56736; AAD14869.1; --  
CC EMBL; AE000290; AAC75050.1; --  
CC EMBL; D90837; BAA15806.1; --  
CC EcoGene; EG14265; nac.  
CC InterPro; IPR000847; --  
CC Pfam; PF00126; HTH\_1; 1.  
CC PRINTS; PR00039; HTHLYSR.  
CC PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.  
CC Transcription regulation; DNA-binding; Activator; Repressor;  
CC Nitrate assimilation.  
CC DNA\_BIND 18 37 H-T-H MOTIF (POTENTIAL).  
CC SEQUENCE 305 AA; 32835 MW; B0B5DC408C6C6821 CRC64;

Query Match 9.8%; Score 142.5; DB 1; Length 305;  
Best Local Similarity 25.3%; Pred. No. 0.00019;  
Matches 75; Conservative 45; Mismatches 78; Indels 99; Gaps 14;  
QY 1 MNPDIQLTLLSIIDEGSFEGASIALSISPSAVSQRVKALEHHVGRVLSRT-OPAKATEA 59  
DB 1 MNPFRLLKYFKYKIDIGSLTQAAEVLHIAQALPSQQVATLEGEINQQLLIRTKRGVTPDA 60  
QY 60 GEVLVQAARKWVLQAEKTAQ-----LSGRLA-----EPLTIATNADS 98  
DB 61 GKLYTHAR-AIRQCE-QQLAVHNVGQALSCQVSIIGFAPGTAASITMPLLOAVRAE- 117  
QY 99 LSTWFFPPFNEVASMGATITLRLDEAHTLSLLRRGDVLGAVTREANPVAG----- 150  
DB 118 ----FPFVIYLLHENSAGVNLNEKLIN--HOLDM-----AVIYHSPVAGVSSQALLK 163  
QY 151 -----C--EVELGTMRHLATATSLRDYAMVDGKLDWAAMPVLRFGPKDVLQDR 198  
DB 164 EDLFLVGTQDCPGQSDVNAIAQMNLFPS-----DYSAIRL----- 200  
QY 199 DLDRYDGPVGRBRVSISSAEFGF-----AIRRGLGWGLLPETQAAAPMLKA 246  
DB 201 ----RVDEAFSLRL----TAKVIGEIESITATLTAIAASGMGAVLPESAARSLOGA 249

RESULT 8  
ID NAC\_KLEAE STANDARD; PRT; 305 AA.  
AC Q08597;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE NITROGEN ASSIMILATION REGULATORY PROTEIN NAC (NITROGEN ASSIMILATION  
DE CONTROL PROTEIN).  
GN NAC.  
OS Klebsiella aerogenes.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Klebsiella.  
OX NCBI\_TaxID=28451;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=W70 / KC1043;  
RX MEDLINE=93209957; PubMed=8458853;  
RA Schwacha A., Bender R.A.;  
RT "The nac (nitrogen assimilation control) gene from Klebsiella  
RT aerogenes.";

RL J. Bacteriol. 175:2107-2115(1993).  
CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR FOR THE HUT, PUT AND URE  
CC OPERONS AND REPRESSOR FOR THE GDH AND GLUT OPERONS IN RESPONSE TO  
CC NITROGEN LIMITATION. NEGATIVE REGULATOR OF ITS OWN EXPRESSION.  
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L01114; AAA18173.1; --  
CC EMBL; B47099; B47099.  
CC InterPro; IPR000847; --  
CC Pfam; PF00126; HTH\_1; 1.  
CC PRINTS; PR00039; HTHLYSR.  
CC PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.  
CC Transcription regulation; DNA-binding; Activator; Repressor;  
CC Nitrate assimilation.  
CC DNA\_BIND 18 37 H-T-H MOTIF (POTENTIAL).  
CC SEQUENCE 305 AA; 32754 MW; 775F219AF32AFADC CRC64;

Query Match 9.6%; Score 140.5; DB 1; Length 305;  
Best Local Similarity 28.1%; Pred. No. 0.00028;  
Matches 77; Conservative 39; Mismatches 105; Indels 53; Gaps 12;  
QY 1 MNPDIQLTLLSIIDEGSFEGASIALSISPSAVSQRVKALEHHVGRVLSRT-OPAKATEA 59  
DB 1 MNPFRLLKYFKYKIDIGSLTQAAEVLHIAQALPSQQVATLEGEINQQLLIRTKRGVTPTEA 60  
QY 60 GEVLVQAARKWVLQAE-----TKAOLSGRLAEIPLTIATNADSLSTWPPVF 107  
DB 61 GKLYTHAR-TILRQCEQAQLAVNVGQTLRGQVSIIGLAPGTAASAITWPLLOT-----VR 115  
QY 108 NEVASMGATITLRLDEAHTL--SLLRGDVLGAVTREANPVAGVEVGLGTMRLHIA 165  
DB 116 NELPE-----VMVYLOESSCTALNDKLLAGQLDMAVLYERSPVAG-----IVS 158  
QY 166 TPSLR-DAYMVDGK-----LDWAAMPVLR-FGPKDVLQDRDLGRVDGPGVRRRSI-- 215  
DB 159 QPLLKEDLVLTGTRDCPGQSDVLTAVAEMLNLFPRDYSAVR---ARVTEAFTLRRLSAKI 215  
QY 216 ---VPSAEGFGEAIRRGLGWGLLPETQAAAPMLKA 246  
DB 216 IGEIESITITLTAIAASGMGATVLPESAARSLOGA 249

RESULT 9  
ID ILVY\_ECOLI STANDARD; PRT; 297 AA.  
AC P05827;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE TRANSCRIPTIONAL ACTIVATOR PROTEIN ILVY.  
GN ILVY.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=86111952; PubMed=3003115;  
RA Wek R.C., Hatfield G.W.;  
RT "Nucleotide sequence and in vivo expression of the ilvY and ilvC  
RT genes in Escherichia coli K12. Transcription from divergent  
RT overlapping promoters.";

J. Biol. Chem. 261:2441-2450(1986).

RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-K12 / MG1655;  
 RX MEDLINE-92358234; PubMed-1379743;  
 RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;  
 RT "Analysis of the *Escherichia coli* genome: DNA sequence of the region  
 from 84.5 to 86.5 minutes.";  
 RL Science 257:771-778(1992).

RL [3]  
 RN SEQUENCE OF 264-297 FROM N.A.  
 RP STRAIN-K12;  
 RX MEDLINE-8717471; PubMed-3550695;  
 RA Lawther R.P., Wek R.C., Lopes J.M., Pereira R., Taillon B.E.,  
 RA Hatfield G.W.;  
 RT "The complete nucleotide sequence of the *ilvGMDA* operon of  
*Escherichia coli* K-12.";  
 RL Nucleic Acids Res. 15:2137-2155(1987).

CC -!- FUNCTION: THIS PROTEIN ACTIVATES THE TRANSCRIPTION OF THE *ilvC*  
 GENE IN THE PRESENCE OF ACETOLACTATE OR ACETOHYDROXYBUTYRATE.  
 CC *ilvC* IS ALSO A NEGATIVE REGULATOR OF ITS OWN EXPRESSION.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO THE *LYSR* FAMILY OF TRANSCRIPTIONAL  
 REGULATORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M11689; AAA24028.1; -  
 DR EMBL; M87049; AAA67576.1; -  
 DR EMBL; AF000453; AAC77493.1; -  
 DR EMBL; M32253; AAA24025.1; -  
 DR PIR; B26287; RGEICY.  
 DR PIR; S30671; S30671.  
 DR EcoGene; EG10503; *ilvC*.  
 DR InterPro; IPR000847; -  
 DR Pfam; PF00126; HTH\_1; 1.  
 DR PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.  
 DR Branched-chain amino acid biosynthesis; Transcription regulation;  
 KW Activator; Repressor; DNA-binding.  
 FT DNA\_BIND 18 37 H-T-H MOTIF (BY SIMILARITY).  
 SQ SEQUENCE 297 AA; 33204 MW; E8FA6A0FC52D3815 CRC64;

Query Match 9.6%; Score 139.5; DB 1; Length 297;  
 Best Local Similarity 23.3%; Pred. NO. 0.00032;  
 Matches 67; Conservative 50; Mismatches 115; Indels 55; Gaps 13;

QY 6 LDTLSIIDEGSFEGASLALSISSAVSORVKALEHHVGRVLVSR-TOPAKATEAGEVLV 64  
 DB 6 LKTFHLAESRFRGKSARAHMVSTLRSQIRLEDLGQPLFVRDNRVTTLTEAGEELR 65  
 QY 65 QAARKWVLLQAEFTK-----AQLSGRLAEPLTAINADSLSTWFPVFNVEV-ASWGGA 116  
 DB 66 VPAQOTLLQYQOLRHTIDQGPSLSGE-----LHIFCSVTAAYSHLPLPILDRFRAEHPV 120  
 QY 117 TLTLEDEAHTLSLLRGDVLGAVTRRANPVAGCEVVELGTMRHLAIA--TPSL-----R 170  
 DB 121 EIKLTGDAADAMEKVVVGEADLAIGKPTLPG--AVAFSMLNLAIVLVIAPALPCVPR 178  
 QY 171 DAYMVDGKLDWAAMPVLRPGKVDLDRDLDRVDPGVRERVSI----- 215  
 DB 179 NOVSVF-PPDSTVPF-----INADQ-----GPV-RRRIELWFRNNKISNPMIYAT 222  
 QY 216 VPSAGFGFETRRGLGWLLPPT--QAAPMLKAGVILLDETPIDTP 260  
 DB 223 VGGHEAMYSVALGCGVALLPEVVLNENSPPEVPRNRYMILERSDEKTP 269

RESULT 10  
 METR\_HAEIN STANDARD; PRT; 309 AA.  
 AC P45349;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE TRANSCRIPTIONAL ACTIVATOR PROTEIN METR.  
 GN METR OR H11739.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RD / KW20 / ATCC 51907;  
 RX MEDLINE-95350630; PubMed-7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of *Haemophilus*  
*influenzae* Rd.";  
 RL Science 269:496-512(1995).

CC -!- FUNCTION: CONTROL OF THE LAST STEP IN METHIONINE BIOSYNTHESIS;  
 METR IS A POSITIVE ACTIVATOR OF THE *MET*, *MET* AND *MET* GENES.  
 CC METR IS ALSO A NEGATIVE REGULATOR OF ITS OWN EXPRESSION  
 (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE *LYSR* FAMILY OF TRANSCRIPTIONAL  
 REGULATORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U32846; AAC23383.1; -  
 DR TIGR; H11739; -  
 DR InterPro; IPR000847; -  
 DR Pfam; PF00126; HTH\_1; 1.  
 DR PRINTS; PR00039; HTH\_LYSR.  
 DR PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.  
 KW Transcription regulation; DNA-binding; Methionine biosynthesis;  
 KW Activator; Repressor.  
 FT DNA\_BIND 23 42 H-T-H MOTIF (POTENTIAL).  
 SQ SEQUENCE 309 AA; 35344 MW; AOC2F0277DE8B274 CRC64;

Query Match 9.5%; Score 139; DB 1; Length 309;  
 Best Local Similarity 26.5%; Pred. NO. 0.00037;  
 Matches 72; Conservative 45; Mismatches 105; Indels 50; Gaps 15;

QY 6 LDTLSIIDEGSFEGASLALSISSAVSORVKALEHHVGRVLVSR-TOPAKATEAGEVLV 64  
 DB 11 LKTLALAKETGSVSLAAKRVLTQSALSHQIKLIEQFGLPLFERKSNPLRFTSAGERLI 70  
 QY 65 QAARKWV--LQAEYKAOLSGRLAEI-----PLTIAINADSLSTWFPVFNVEAS-RGG 115  
 DB 71 RLANEVMPKVIDAERD-----LARVKHGDAGQLRIAVECHTCFDWLMPAMDEFFQHG- 123  
 QY 116 ATTLRLDEAHT---LSLLRRGDLGAVTREA-----NPVAGCEVVELGTMRHLA 163  
 DB 124 -LVLDIVSGFHTDPVGLLSHRAD--WAIVSEIHNDVDVIFKPLFSYEMGICSKNH-- 178



```

QY 164 IATPSLRDAYMVDGKLD--WAAMPVLRFPGPKDVLQDRDLGRV---DGPVGRRRVSVIPS 218
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 --SLAEKDINEAFIDETWYTPV---PDDML---DLWRKVLKSKGINPTRRTTTLTI 229
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 219 AEGFGEAIRRGGLGWLGPETQAAAPMLKAGEVI 250
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 AMIQLVASRRGI--ATIPYWAALPYLEKGYV 259
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
YF_ER_ECOLI
ID YF_ER_ECOLI STANDARD; PRT; 308 AA.
AC P77500;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN XAPA-LIG INTERGENIC REGION.
GN YF_ER.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000328; AAC75462.1; -
CC EMBL; D90870; BAA16280.1; -
CC EcoGene; EG14159; yfer.
CC InterPro; IPR000847; -
CC Pfam; PF00126; HTH_1; 1.
CC PRINTS; PR00039; HTHLYSR.
CC PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW Hypothetical protein; Transcription regulation; DNA-binding.
FT DNA_BIND 20 39 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 308 AA; 33903 MW; 23B6A771E4AB587 CRC64;

Query Match 9.5%; Score 138.5; DB 1; Length 308;
Best Local Similarity 25.1%; Pred. No. 0.0004;
Matches 78; Conservative 51; Mismatches 119; Indels 63; Gaps 18;

```

```

QY 5 QLTLLSIIDEGSEFASLALISPSAVSORVKALEHHVG-RVLVSTQPAKATEADEVL 63
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7 QLKVFVVAQEKSFAGERIGLSQSAVSHSVKLEHNHTGVRLDRTTREVLTDAQ-- 64
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 VQAARKVLLQAEKTA-----OLSGRLAEIPLTIALNADSLSTWFP--VFNEVA 111
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 -QLALRLERLDELNLTDRGTGRMGQQLSGKVR-----VAASQTISAHLIPOCIAESHR 117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 112 SWGGATLTLRLEDEAHTLSLLRRGDVLGAVTREANPVAG--CEVVELGTMRLHAIATPSL 169
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 RYPDIQFVLHDPQOQWVNESIRQGDVDFGVIDPGVGDLOCEAI-----LSEPF 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 ----RD--AYMVDGKLDWAAMPVLRFPGPKDVLQDRDLGR--VDGPVGR--RVSIVPSAE 220
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 LLCHRSALAVEDYVPWQALQ---GAKLVLDQYASGRPLIDAALARNGTQANIV---Q 221
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 221 GFGE-----AIRGLGWLGPETQAAAPMLKAGEVILLDEIPIDTPMYWGRW---RLES 271
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 EIGHPATLFPMAAGIGISILP-ALALP-LPEGSPLVVKRI---TPVVERQLMLVLRKNR 276
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 272 SLARLTDAVVD 282
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 SLSTAAREALWD 287

RESULT 12
METR_SALTY
ID METR_SALTY STANDARD; PRT; 317 AA.
AC P05984;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TRANSCRIPTIONAL ACTIVATOR PROTEIN METR.
GN METR.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87307964; PubMed=3040668;
RA Plamann L.S., Stauffer G.V.;
RT "Nucleotide sequence of the Salmonella typhimurium metR gene and the
RT metR-metE control region."
RL J. Bacteriol. 169:3932-3937(1987).
RN [2]
RP REVISIONS.
RA Urbanowski M.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CONTROL OF THE LAST STEP IN METHIONINE BIOSYNTHESIS;
CC METR IS A POSITIVE ACTIVATOR OF THE MET, METE AND METH GENES.
CC METR IS ALSO A NEGATIVE REGULATOR OF ITS OWN EXPRESSION.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M17356; AAB88753.1; -
CC StyGene; SG10227; metr.
CC InterPro; IPR000847; -
CC Pfam; PF00126; HTH_1; 1.
CC PRINTS; PR00039; HTHLYSR.
CC PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW Transcription regulation; DNA-binding; Methionine biosynthesis;
KW Activator; Repressor.

```

FT DNA\_BIND 19 38 H-T-H MOTIF (BY SIMILARITY).  
SQ SEQUENCE 317 AA: 35585 MW; 90CD503F84BE8CC4 CRC64;

Query Match 9.3%; Score 136.5; DB 1; Length 317;  
Best Local Similarity 23.9%; Pred. No. 0.0006;  
Matches 69; Conservative 29; Mismatches 110; Indels 81; Gaps 8;

QY 6 LDTLLSIIDEGSFGASLISPSAVSRVKALEHHVG-RVLVSTQPAKATEACEVLV 64  
DB 7 LKTLQALNRSGSLAAAVLHQTSALSQHSQFSDLEORLGRFLFRVRSQBLRFTPOGEVLV 66  
QY 65 QAARKMVLLOAETKAQLSGRLEAIPLTIAINADSLSTWTFPPV-NEVASW-----GG 115  
DB 67 QLANO-VLPQISRALQACNEPOQTRURIAIECHSCIOWLTPALENFRASWPQVEMDFTS 125  
QY 116 ATLTLRLDEAHTLSLLRGDVLGAVTRANPVAGEV----- 153  
DB 126 VTFDQPALQOGELDVLVMTSDILPRSELHYSNMFDFEVLRLVLPADHPLASKTQITPEDLA 185  
QY 154 -----VELGTMRHL--AIATPSLRDA-----YVVDGKLDWAAMPVLRFGPK 192  
DB 186 SETLLIYPQRSRLDWRHFLQAPGISPLKSVNTLLLIQWVAARMGTAALPHW----- 240  
QY 193 DVLQDRDLGRVDGPVGRRRVSIVPSAEGFGEAIRRGGLGWLPPETQAA 241  
DB 241 -----VVESVERQGLVVTTLGDGLWSRLYAA 267

RESULT 13  
LRRR\_SYPN7  
ID LRRR\_SYPN7 STANDARD; PRT; 294 AA.  
AC P52691;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE PROBABLE TRANSCRIPTIONAL REGULATOR LRRR.  
GN LRRR.  
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=1140;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96281517; PubMed=8661945;  
RA Anandan S., Nalty M.S., Cogdell D.E., Golden S.S.;  
RT "Identification of two classes of transcriptional regulator genes in  
the cyanobacterium Synechococcus sp. strain PCC 7942.";  
RL Arch. Microbiol. 166:58-63(1996).  
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; LA1665; AAB38541.1;  
DR InterPro; IPR000847;  
DR Pfam; PF00126; HTH\_1;  
DR PRINTS; PR00039; HTHLYSR.  
DR PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.  
KW Transcription regulation; DNA-binding.  
FT DNA\_BIND 18 37 H-T-H MOTIF (POTENTIAL).  
SQ SEQUENCE 294 AA: 32068 MW; F9738747F149E335 CRC64;

Query Match 9.3%; Score 136; DB 1; Length 294;  
Best Local Similarity 23.4%; Pred. No. 0.00061;  
Matches 68; Conservative 42; Mismatches 87; Indels 94; Gaps 12;

QY 1 MNPIQDITLLSIIDEGSFGASLISPSAVSRVKALEHHVGRVLSRTQ-PAKATEA 59  
DB 1 MNITQIQIFAAVETGTFSAALQLDLSQSAVSRAATALEDELGVLLSGRGARPTRV 60  
QY 60 GEVLVQAARKMVLLOAET--KAQLSGRLEAIPLTIAINADSLSTWTFPP--VFNEVASWG 114  
DB 61 GERVLQAQMLQLHDSIVHEVNLKGLQGHRLRIASFRSAATHVLPPRLALFRQRCP-- 118  
QY 115 GATLTLRLDEAHTLSLLRRGDV-LG-----AVTRE-----ANPV--- 148  
DB 119 GVSVSIIETDPQGVQEQALREGKVDIGLLPLPRSEEDTWEITRDEVVLLPSTAHPLGQP 178  
QY 149 -----ACCEVVELGTMRHLAIATPSLRDAYMYVDGKLDWAAMPVLRFGPK 192  
DB 179 LTWAQLSRVEFTILYNAECTT---AVRQHWATARQELKVAY----- 216  
QY 193 DVLQDRDLGRVDGPVGRRRVSIVPSAEGFGEAIRRGGLGWLPPETQAA 243  
DB 217 ETKEDSTI-----VSMV-----AAGLGAAILPRLAAVPI 245

RESULT 14  
METR\_ECOLI  
ID METR\_ECOLI STANDARD; PRT; 317 AA.  
AC P19797;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE TRANSCRIPTIONAL ACTIVATOR PROTEIN METR.  
GN METR.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=K12;  
RX MEDLINE=90384950; PubMed=2205852;  
RA Maxon M.E., Wigboldus J., Brot N., Weissbach H.;  
RT "Structure-function studies on Escherichia coli MetR protein, a  
putative prokaryotic leucine zipper protein.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:7076-7079(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=K12 / MG1655;  
RX MEDLINE=92358234; PubMed=1379743;  
RA Daniels D.L., Plunkett G.III, Burland V.D., Blattner F.R.;  
RT "Analysis of the Escherichia coli genome: DNA sequence of the region  
from 84.5 to 86.5 minutes.";  
RL Science 257:771-778(1992).  
RN [3]  
RP SEQUENCE OF 1-19 FROM N.A.  
RX MEDLINE=89098936; PubMed=2643109;  
RA Maxon M.E., Redfield B., Cai X.-Y., Shoeman R., Fujita K., Fisher W.,  
RA Stauffer G., Weissbach H., Brot N.;  
RT "Regulation of methionine synthesis in Escherichia coli: effect of  
the MetR protein on the expression of the metE and metR genes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:85-89(1989).  
CC -!- FUNCTION: CONTROL OF THE LAST STEP IN METHIONINE BIOSYNTHESIS;  
METR IS A POSITIVE ACTIVATOR OF THE META, METE AND METH GENES.  
CC METR IS ALSO A NEGATIVE REGULATOR OF ITS OWN EXPRESSION. BINDS  
HOMOCYSTEINE AS AN INDUCER.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC EMBL; M37630; AAA62781.1; -.
DR EMBL; M87049; AAA67624.1; -.
DR EMBL; AE00458; AAC76831.1; -.
DR EMBL; J04155; AAA24159.1; -.
DR PIR; A36066; A36066
DR PIR; S30718; S30718.
DR Ecogene; EG10591; metr.
DR InterPro; IPR000847; -.
DR Pfam; PF00126; HTH_1; 1.
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW Transcription regulation; DNA-binding; Methionine biosynthesis;
KW Activator; Repressor.
FT DNA_BIND 19 38 H-T-H MOTIF (BY SIMILARITY).
FT CONFLICT 19 19 L -> T (IN REF. 3).
SQ SEQUENCE 317 AA; 35629 MW; 513F5C710803800D CRC64;

Query Match 8.9%; Score 129.5; DB 1; Length 317;
Best Local Similarity 23.3%; Pred. No. 0.0022;
Matches 69; Conservative 33; Mismatches 111; Indels 83; Gaps 9;

QY 6 LDTLLSIIDEGSPEGASLALSISPSAVSQRKALEHHVG-RVLVSRTPAKATEAGEVLV 64
DB 7 LKTLQNRNGSLAAATLHQTQSALSHQFSDLEQRLGFRLFVRKSQPLRFTPGGEILL 66

QY 65 QAARKMVLQA-----ETKALSGRLAEIPLTIAINADSLSTWFPVNEVASW-----GG 115
DB 67 QLANQ-VLPQISQALQACNEPQOTRLAIAIECHSCIQWLTPALENFHKNPQVEMDFKSG 125

QY 116 ATLRLDEAHTLSLLRRGDVLGAVTNEANPVAGCEV-----153
DB 126 VTEDPQALQGGELDLVMTSDILPRSLGHYSPMFDYEVRLVLADPHPLAAKTRITPEDLA 185

QY 154 -----VELGTRHL---AIATPSLRDA-----YMDGKLDWAMPVLRGPK 192
DB 186 SEYLLIYPVORSRLDWRHFLQAPAGVSFLKSVNTLLILQMVAAARMGIAALPHW-----240

QY 193 DVLQDRDLGVDGPGVRRVRSIVPSAEGFEAIRRGLGGLLPETQAAPMLKAGE 248
DB 241 -----VVESERQGLVVTKTGEGLSRLYAA--VRDGE 272

RESULT 15
TFDR_ALCEU STANDARD; PRT; 295 AA.
AC P10086; P42429;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TRANSCRIPTIONAL REGULATORY PROTEIN TDFR/TFDS.
GN TFDR AND (TFDS OR TFDO).
OS Alcaligenes eutrophus (Ralstonia eutropha).
OG plasmid pJP4
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A. (TFDR AND TFDS).
RC STRAIN=JMP134;
RX MEDLINE=94209236; PubMed=8157603;
RA Matrubutham U., Harker A.R.;
RT "Analysis of duplicated gene sequences associated with tfdr and tfds
in Alcaligenes eutrophus JMP134";
RL J. Bacteriol. 176:2348-2353(1994).
RN [2]
RP SEQUENCE OF 1-180 FROM N.A. (TFDS).
RC STRAIN=JMP134;
RX MEDLINE=87250253; PubMed=3036764;
RA Streiber W.R., Timmis K.N., Zenk M.H.;
RT "Analysis, cloning, and high-level expression of 2,4-
dichlorophenoxyacetate monooxygenase gene tfda of Alcaligenes

```

```

entrophus JMP134.";
RT J. Bacteriol. 169:2950-2955(1987).
RN [3]
RX IDENTIFICATION OF PROTEIN, AND POSSIBLE DNA-BINDING REGION.
RX MEDLINE=88320486; PubMed=3413113;
RA Henikoff S., Haughn G.W., Calvo J.M., Wallace J.C.;
RT "A large family of bacterial activator proteins.";
CC Proc. Natl. Acad. Sci. U.S.A. 85:6602-6606(1988).
CC -1- FUNCTION: INVOLVED IN REGULATION OF 3-CHLOROCATECHOL DEGRADATION.
CC TRANSCRIPTIONAL REGULATOR OF TDFB EXPRESSION. ACTS AS A REPRESSOR
CC IN THE ABSENCE OF ITS EFFECTOR (EITHER 2-CIS-CHLORODIENE LACTONE
CC OR CHLOROMALEYLACETATE) BUT ACTS AS AN ACTIVATOR WHEN ITS EFFECTOR
CC IS PRESENT.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- MISCELLANEOUS: THE PRODUCTS OF THE DUPLICATED GENE TFDR AND TFDS
CC ARE IDENTICAL.
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M98445; AAA5063.1; -.
CC EMBL; M16730; -. NOT ANNOTATED_CDS.
CC InterPro; IPR000847; -.
CC Pfam; PF00126; HTH_1; 1.
CC PRINTS; PR00039; HTHLYSR.
CC PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW Aromatic hydrocarbons catabolism; Transcription regulation;
KW DNA-binding; Activator; Repressor; Plasmid.
FT DNA_BIND 18 37 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 295 AA; 32070 MW; C6C477BE8127A41A CRC64;

Query Match 8.8%; Score 128; DB 1; Length 295;
Best Local Similarity 24.4%; Pred. No. 0.0027;
Matches 76; Conservative 45; Mismatches 138; Indels 52; Gaps 12;

QY 5 QLTLLSLIIDEGSPEGASLALSISPSAVSQRKALEHHVGRLVLSRT-OPAKATEAGEVL 63
DB 5 QLRFYVAAAEENGVGAARLRHLISQPPVTRQIHAEQHLGVLLFERSARGVQLTPAGAAF 64

QY 64 VQARKMVLQA-----ETKALSGRLAEIPLTIAINADSLSTWFPVNE-VASWGATL 118
DB 65 LEDARMLELGRTSVDRSRAASRGEIQ--LDIGYLGTAIVQVTPALLHAFTQAVPGATL 122

QY 119 TLRLDEAHTLSLLRRGDVLGAVTNEANPVAGCEVVELGTMRLHAIATPSLRDAYMVDGK 178
DB 123 SLALMPKVRQIEALRAGTIHILGVRFPQEPGITVEHLHYERLYIAAGSSTARQLRQDPT 182

QY 179 LDWAMPVLRFGPKDVLQDRDLGR---VDGPVG-RRRSIVPSAEGFE-----AIR 227
DB 183 LLRLKESLVLFPK-----EGRPSFADEVIALMRAGVPRVTAIVEDVNAALGLVA 234

QY 228 RGLGWGLLPETQAA---PMLKAGEVI-LLDEIPIDTPTMYQWRLESRSRLARLTD-----278
DB 235 AGAGTVLPASVAAIRRPFRVTMWMADASKVPV-----SLTYLTDSRPV 280

QY 279 --AVVDAAIEG 287
DB 281 LRAFLDVAREG 291

Search completed: September 13, 2001, 17:57:11
Job time: 296 sec

```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:56:48 ; Search time 37.95 Seconds  
(without alignments)  
1011.027 Million cell updates/sec

Title: US-09-105-117i-3  
Perfect score: 1460  
Sequence: 1 MNPIQLDRLLSIDGSEF.....RSLARLTDAVVDAIEGLRP 290

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL16:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_unclassified:\*  
13: sp\_vertebrate:\*  
14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	590.5	40.4	300	2 Q9K4K7	Q9K4K7 streptomyc
2	472.5	32.4	303	2 O87968	O87968 streptomyc
3	410	28.1	300	2 O9HW38	O9HW38 pseudomonas
4	361.5	24.8	298	2 Q9KUN3	Q9KUN3 vibrio chol
5	177	12.1	297	2 Q9HX26	Q9HX26 pseudomonas
6	175	12.0	300	2 Q9IAJ4	Q9IAJ4 pseudomonas
7	174	11.9	295	2 Q9I085	Q9I085 pseudomonas
8	170	11.6	310	2 Q9I122	Q9I122 pseudomonas
9	163.5	11.2	286	2 Q9IA35	Q9IA35 pseudomonas
10	156	10.7	304	2 Q9I4D7	Q9I4D7 pseudomonas
11	152.5	10.4	292	2 Q9I0Y5	Q9I0Y5 pseudomonas
12	152.5	10.4	295	2 Q9KM14	Q9KM14 vibrio chol
13	151	10.3	302	2 Q9KNH9	Q9KNH9 vibrio chol
14	149.5	10.2	339	2 Q9HZ25	Q9HZ25 pseudomonas
15	149	10.2	303	2 Q9I1E4	Q9I1E4 pseudomonas
16	144	9.9	278	2 O86059	O86059 azospirillu
17	143.5	9.8	297	2 Q9L6S5	Q9L6S5 s 98% ident
18	142	9.7	342	2 O67999	O67999 bradyrhizob
19	141.5	9.7	311	2 Q9HXL7	Q9HXL7 pseudomonas

## ALIGNMENTS

RESULT	ID	Q9K4K7	PRELIMINARY;	PRT;	300 AA.
AC	O9K4K7				
DT	01-OCT-2000	(Tremblrel. 15, Created)			
DT	01-OCT-2000	(Tremblrel. 15, Last sequence update)			
DT	01-MAR-2001	(Tremblrel. 16, Last annotation update)			
DE	PUTATIVE LYSR-FAMILY TRANSCRIPTIONAL REGULATOR.				
GN	SC5F8.17C.				
OS	Streptomyces coelicolor.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.				
OX	NCBI_taxid=1902;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=A3(2);				
RA	Seeger K.J., Harris D.;				
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=A3(2);				
RA	Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;				
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=A3(2);				
RA	MEDLINE=97000351; PubMed=843436;				
RA	Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.;				
RA	Kinashi H., Hopwood D.A.;				
RT	"A set of ordered cosmids and a detailed genetic and physical map for				
RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";				
RL	Mol. Microbiol. 21:77-96(1996).				
CC	- - SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.				
DR	EMBL: AL357613; CAB93745.1; -				
DR	InterPro: IPR000847; -				
DR	InterPro: IPR001993; -				
DR	Pfam: PF00126; HTH_1; 1.				
DR	PRINTS; PR00039; HTHLYSR.				
DR	PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.				
DR	PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.				
KW	DNA-binding; Transcription regulation.				
SQ	SEQUENCE 300 AA; 32205 MW; 7122FE5A539D642B CRC64;				

Q9i6x3 pseudomonas  
Q9krd6 vibrio chol  
O86805 streptomyc  
P96725 bacillus su  
Q9hu66 pseudomonas  
O52827 rhizobium l  
Q9i568 pseudomonas  
Q9i6n2 salmonella  
Q9i0f7 pseudomonas  
O66393 azorhizobiu  
Q9rdh0 streptomyc  
Q9l3r4 anabaena sp  
Q9jwg7 neisseria m  
Q9rcy3 streptomyc  
Q9fi86 alcaligenes  
Q9jxg8 neisseria m  
Q9i569 pseudomonas  
Q9i4f4 pseudomonas  
Q9hxb8 pseudomonas  
Q9i0n6 pseudomonas  
Q9zbw3 streptomyc  
Q9klv6 vibrio chol  
Q9i709 pseudomonas  
O30441 bordetella  
Q9rh21 zymomonas m  
Q9hy39 pseudomonas

20 140 9.6 312 2 Q9i6x3  
21 138.5 9.5 305 2 Q9krd6  
22 138.5 9.5 306 2 O86805  
23 138 9.5 293 2 P96725  
24 136 9.3 294 2 Q9HU66  
25 136 9.3 298 2 O52827  
26 136 9.3 298 2 Q9i568  
27 135.5 9.3 317 2 Q9L6N2  
28 135 9.2 309 2 Q9i0F7  
29 134 9.2 311 2 O66393  
30 134 9.2 313 2 Q9RDH0  
31 133.5 9.1 312 2 Q9L3R4  
32 133 9.1 309 2 Q9JWG7  
33 132 9.0 278 2 Q9RCY3  
34 132 9.0 307 2 Q9FI86  
35 131 9.0 309 2 Q9JXG8  
36 131 9.0 314 2 Q9I569  
37 130 8.9 296 2 Q9I4F4  
38 130 8.9 317 2 Q9HXB8  
39 129.5 8.9 298 2 Q9I0N6  
40 129 8.8 292 2 Q9ZBW3  
41 128.5 8.8 313 2 Q9KLv6  
42 128 8.8 302 2 Q9I709  
43 127.5 8.7 291 2 O30441  
44 127.5 8.7 302 2 Q9RH21  
45 127 8.7 306 2 Q9HY39

```
Query Match 40.4%; Score 590.5; DB 2; Length 300;
Best Local Similarity 44.5%; Pred. No. 1.6e-37;
Matches 129; Conservative 51; Mismatches 103; Indels 7; Gaps 5;

QY 5 QDITLSIIIDEGSFEGASLALISPSAVSQRVKALEHHVGRVLSRTQPAKATEAGEVLV 64
D 12 QVETLLAVDEGTFDAAAALHWTSPAVSQRVKALEQRTGRVLLLTQKPVTRATDSGAVLV 71
QY 65 QAARKVLIQAEKKAOLSRLEIP--LTIATNADSLSTWTFPPVNEVASWGATLTLEL 122
D 72 RLARQVARLERDASAEGLRGECEPTRVTAVNADSLATWFLPALTRIPREPALCFELRR 131
QY 123 EDEAHTLSLLRGDVLGAVTREANPVAGCEVVELCTMRHLATATPSLRDAYWVDGKL--D 180
D 132 EDEGHTATLLRGVNVAAVTSSEPPVPGCTVTLGRMYLPCCAPEFA--ARHLDAPLRET 190
QY 181 WAAMPVLRGPKDVLQDRDLGRVDGPGVRRRVSIVPSAEGFGEAIRRGLGWLGLLPETQA 240
D 191 VAGAPVVVFDRRDDFQDSFARRLGHGGAAR--HYVPTSEGEFVEAAAGLGMVGPQQA 249
QY 241 APMKAGEVI--LLEIDPIDTPMYQWRLESRLARLTDAVVDAAIEGLR 289
D 250 EPLLRTGLVTFAPDLAVDVTLYQWQMKLDSPALATVADAVVTAADALR 299

RESULT 2
ID O87968 PRELIMINARY; PRT; 303 AA.
AC O87968;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ORF5.
OS Streptomyces griseus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
RN [1]
RP SEQUENCE FROM N.A.
RA Uneyama T.;
RT "Open reading frame encoded around afsA gene.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.
DR EMBL; AB011413; BAA32133.1;
DR InterPro; IPR000847;
DR InterPro; IPR001993;
DR Pfam; PF00126; HTH_1;
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 303 AA; 31866 MW; FID146DD95AAOCE3 CRC64;

Query Match 32.4%; Score 472.5; DB 2; Length 303;
Best Local Similarity 39.3%; Pred. No. 1.7e-28;
Matches 125; Conservative 49; Mismatches 93; Indels 51; Gaps 12;

QY 1 MNPIDLD---TLLSIIDEGSFEGASLALISPSAVSQRVKALEHHVGRVLSRTQPAKAT 57
D 1 MSELPLDQVRLTLLAVDEGTFDAAAALHWTSPAVSQRVKALEQRTGRVLLMTKPVRT 60
QY 58 EAGEVLQAARKVLIQAEKKAOL--SGRLAIEPLTIATNADSLSTWTFPPVNEVASWG 115
D 61 ESSEV--VALARLARLEAHEAQAALGMSGGEPTELLPIAVNSDLATW-----SWQP 110
QY 116 ATLTLRLEDAHTL-----SLLRRGD--VLGAVTREANPVAGCEVVELGTMR 160
D 111 CGACRRSMDSATTPGPGPYRPAAGGAGDGRDLVAGGDLGFGPA-----GMR 163
QY 161 HLAIATPSLRDAYWVDGKLDWAAM-----PVLRFQPKDVLQDRDLGRVDGPGVR--R 211
```

```
Db 164 YLPVSPGSPADRWL--GRDGTALRELIGCAPVWCFDRDDLDQAFV--RRLGPGARPSA 219
QY 212 RVSIVPSAEGFGEATRRGLGWLGPETQAPMLKAGEVI--LLEIDPIDTPMYQWRLES 270
D 220 RRHLVPTSEGFANAVASGNGMVGMPVEQAEPLLSDGRLVRLAPEPTVDVPLVWQWKLES 279
QY 271 RSLARLTDAVVDAAIEGL 288
D 280 PALAAVAEAAVAEAAEAL 297

RESULT 3
ID Q9HW38 PRELIMINARY; PRT; 300 AA.
AC Q9HW38;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE INHIBITOR OF CHROMOSOME INITIATION ICIA.
GN ICIA OR PA4363.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K.; Pham X.-Q.T.; Erwin A.L.; Mizoguchi S.D.; Warrenner P.;
RA Hickey M.J.; Brinkman F.S.D.; Hufnagle W.O.; Kowalik D.J.; Lagrou M.;
RA Garber R.L.; Goltner L.; Tolentino E.; Westbrock-Wadman S.; Yuan Y.;
RA Brody L.L.; Coulter S.N.; Folger K.R.; Kas A.; Larbig K.; Lim R.M.;
RA Smith K.A.; Spencer D.H.; Wong G.K.-S.; Wu Z.; Paulsen I.T.;
RA Reizer J.; Sailer M.H.; Hancock R.E.W.; Lory S.; Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964 (2000).
DR EMBL; AE004852; AAG07751.1;
DR InterPro; IPR000847;
DR Pfam; PF00126; HTH_1;
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
SQ SEQUENCE 300 AA; 32451 MW; BFC7F298E51CFBE9 CRC64;

Query Match 28.1%; Score 410; DB 2; Length 300;
Best Local Similarity 36.9%; Pred. No. 1e-23;
Matches 107; Conservative 47; Mismatches 126; Indels 10; Gaps 5;

QY 6 LDTLLSIIDEGSFEGASLALISPSAVSQRVKALEHHVGRVLSRTQPAKATEAGEVLVQ 65
D 9 LAALAAVVEQGGFERAAQAALGLSQSAVSQRIKLEARVGPVLVRETTPHPTDLGRLLN 68
QY 66 AARKMVLIQAEK---AQLSGRLAIEPLTIATNADSLSTWTFPPVNEVASWGATLTLEL 122
D 69 HVQOVRLLGDLQVRWPNLDEGAPERLRIALNADSLATWAAAVGDFCAERRVLLDLYV 128
QY 123 EDEAHTLSLLRGDVLGAVTREANPVAGCEVVELCTMRHLATATPSLRDAYWVDG--KLD 180
D 129 EDOEVLGKMRAGEVAGCVCSARPVAGARSLLGLAMRYGLASPDFIARHPFGRVEAA 188
QY 181 WAAMPVLRGPKDVLQDRDL--DGRVDGPGVRRRVSIVPSAEGFGEAIRRGLGWLGLLPETQ 239
D 189 LAGVPAIVVEQGGFERAAQAALGLSQSAVSQRIKLEARVGPVLVRETTPHPTDLGRLLN 245
QY 240 AAPMLKAGEVI--LLEIDPIDTPMYQWRLESRLARLTDAVVDAAIEGL 288
D 246 VOGELARGELVELPGQVIDVPLVYWHYWRNGELLASLEHLLARAGDGL 295

RESULT 4
Q9KUN3
```

ID	Q9KUN3	PRELIMINARY;	PRT;	298 AA.
AC	Q9KUN3;			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)			
DE	CHROMOSOME INITIATION INHIBITOR.			
GN	VC0482.			
OS	Vibrio cholerae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=666;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EL TOR N16961 / SEROTYPE O1;			
RX	MEDLINE=20406833; Pubmed=10952301;			
RA	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,			
RA	Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,			
RA	Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,			
RA	Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,			
RA	McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,			
RA	Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,			
RA	Fraser C.M.;			
RT	"DNA sequence of both chromosomes of the cholera pathogen Vibrio			
RT	cholerae";			
NC	Nature 406:477-483(2000).			
CC	-!- SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.			
DR	EMBL; AE004134; AAF93655.1; -			
DR	TIGR; VC0482; -			
DR	InterPro: IPR000847; -			
DR	Pfam; PF00126; HTH_1; 1.			
DR	PRINTS; PR00039; HTHLYSR.			
DR	PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.			
KW	DNA-binding; Transcription regulation.			
QY	SEQUENCE 298 AA; 33653 MW; C7776C56A9FD2521 CRC64;			
Query Match 24.8%; Score 361.5; DB 2; Length 298;				
Best Local Similarity 29.9%; Pred. No. 5.1e-20;				
Matches 86; Conservative 65; Mismatches 116; Indels 21; Gaps				
Qy	6	LDTLLSIDSGSPGASLALSTSPSAVSORVKALEHHVGRVLVSTQPAKATEAGEVLVQ	65	
Db	9	IEALDSVVGSGFERAAEQLFITQSASQRIKQLEKLAQPVLIREQPPRPTLVGKKLLG	68	
Qy	66	AARKMWLLQAEKTAQLSGR--LAEIPLTIAINADSLSTWFPFVFNVAWSGATITLRLE	123	
Db	69	LYRRVCLIEQLVPELTNQEHVRPVSMSKSTATNADSLATWLLPALDKYMKSRQVELNLVIY	128	
Qy	124	DEAHTLSLLRRGDVLGAVTREANPGVCEVVELGTMRHLATPSLRDAYMVDGKLDWAA	183	
Db	129	GESRTLDKLNAGEVGAISLEQPTIGCSAEYLGQMEVLCVASPEFYQKYFAKGVTPRSL	188	
Qy	184	M--PVLRFPGKVDLQDRDL-----DGRVDGPVRRRVSVISPSAEGFGAIRRGLGWG	233	
Db	189	IKAPAVSYDQYDELRHKLWDYFAVPRDKVINHTVQ-----SSEAFVRLALSGAAYC	240	
Qy	234	LLPETAAPMLKAGEVI-LLEDTPTDTPMWQRWLESRSRLRLTDVAV	280	
Db	241	LIPRLQIISLESALINPTDFMLSYPIFHHWQLETGVLLLEISAI	288	
RESULT	5			
Q9HX26				
ID	Q9HX26	PRELIMINARY;	PRT;	297 AA.
AC	Q9HX26;			
DT	01-MAR-2001 (TrEMBLrel. 16, Created)			
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)			
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)			
DE	PROBABLE TRANSCRIPTIONAL REGULATOR.			
GN	PA3995			
OS	Pseudomonas aeruginosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OC	Pseudomonas.			
OX	NCBI_TaxID=287;			

Qy	6	LDTLLSIDEGSPGASLALSI	SPSAVSQVRKALEHHVGRVILVSR	T--OPAKATEAGEVLV	64
Db	8	IELFLAVLDLRGSGAARALGRV	SPSVMGTIANLEAEALFALFERTHREARPTALARALA	67	
Qy	65	QAARKWILQAQETKAQISGR	LAETP-----LTAINAD-----	SLSTWPPPVF	107
Db	68	POARMI-----AEQAQ	QOVIALELSQGLSRSLGUVADIDSGRLFGALFTLSERYPLLD	123	
Qy	108	NEVASWGGATLTLRLEDEAHT	LSLLRRGDV---LGAVTREANP-----	VAGCEVVELGT	158
Db	124	VEVLS-----AAQDDA--	LALLHGGRISLCLFCGFAGISVNPQERFQYVGAESLVATIS	173	
Qy	159	MRHLAIATPSLRDAYMYDG	KLDWAAMPVLRFPGKDVLDROD--LDGRVDGPGVRRRYSIVP	217	
Db	174	PRHPALQKPG--QALYL--	-----EELVNVFQIILVASCDDLPLADTRPLIAG	215	
Qy	218	S---AEGFG---EAIRGLG	WGLLPETQAAPMLKAGEVILLD-----EIPIDTPMYQR	265	
Db	216	ACWRDTSLSGTALEMVEAG	ICWGNFPLSRVAPLLATGRVLRLDFRNTKNELKLPVHAIW--	273	
Qy	266	WRLESRSIARLITDAV	V 281		
Db	274	--LKNQPLRKA	AQELV 287		
RESULT	8				
Q91122		PRELIMINARY;	PRT; 310 AA.		
AC	Q91122;				
DT	01-MAR-2001 (Tremblrel. 16, Created)				
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)				
DT	01-MAR-2001 (Tremblrel. 16, Last annotation update)				
GN	PA2123				
OS	Pseudomonas aeruginosa.				
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;				
OC	Pseudomonas.				
OX	NCBI_TaxID=287;				
RN	SEQUENCE FROM N.A.				
RC	STRAIN=PA01.				
RC	MEDLINE=20437337; PubMed=10984043;				
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,				
RA	Hickey M.-J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,				
RA	Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,				
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,				
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,				
RA	Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;				
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an				
RT	opportunistic pathogen."				
RL	Nature 406:959-964(2000).				
DR	EMBL; AE004639; AAG0511.1;				
DR	InterPro; IPR000847;				
DR	Pfam; PF00126; HTH_1;				
DR	PRINTS; PR00039; HTHLYSR.				
SQ	SEQUENCE 310 AA; 34611 MW; 258EL5C239BEFE7C CRC64;				
Query Match	11.6%; Score 170; DB 2; Length 310;				
Best Local Similarity	25.7%; Pred. No. 2.4e-05;				
Matches	77; Conservative 37; Mismatches 98; Indels 88; Gaps				
Qy	5	QDFTLLSIDEGSPGASLALSI	SPSAVSQVRKALEHHVGRVILVSR	TQPAKATEAGEVLV	64
Db	6	EMTTFVSVVDHAHSEFAARR	LGTGTSQVSQRIQOLRRRLGLNLRTPLSLTDPGRTY	65	
Qy	65	QAARKWV--LQAET-----	KAQLSGLAETP-----LTAIN 95		
Db	66	EHA	CRLOEIEQAQASVLDADDLDRGRLQISAPLAFTPYRLAPLLARFAEHFQRLRVDVQ	125	
Qy	96	ADS--LSTWFFPVFVNEVAS	WGGATUTLRLDEAHTLSLLRRGDVLGAVTREANPVAGEV	154	
Db	126	ADDFVNINLP	PRFDALRMG-----RLDDSLV-----ARPT-----	158	



```

DI      01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT      01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE      PROBABLE TRANSCRIPTIONAL REGULATOR.
GN      PAL201.
OS      Pseudomonas aeruginosa.
OC      Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OX      Pseudomonas.
OY      NCBI_TaxID=287;
[1]     RN
RP      SEQUENCE FROM N.A.
RC      STRAIN=PA01.
RX      MEDLINE=20437337; PubMed=10984043;
RA      Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA      Hickey M.J., Brinkman F.S.I., Huffnagle W.O., Kowalik D.J., Lagrou M.,
RA      Garber R.L., Gotlry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA      Brody L.A., Coulter S.H., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA      Smith K.A., Spencer D.N., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA      Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT      "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT      opportunistic pathogen." ;
RL      Nature 406:959-964(2000).
RD      ENBL: AE004550; AAG04590.1; -.
DR      InterPro: IPR000847; -.
DR      Pfam: PF00126; HTH_1; 1.
DR      PRINTS: PR00039; HTLXSR.
DR      PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN1.
SQ      SEQUENCE   304 AA;  33974 MW;  9AA2E24AE5CA8B1 CRC64;

Query Match          10.7%    Score 156:       DB 2: Length 304;
Best Local Similarity 26.7%; Pred.No. 0.00028;
Matches              76; Conservative                    52; Mismatches 109; Indels         48; Gaps

Qy        6 LDLTLSIIDGSEFASIALISPSAVSORVKALEHHVG-RVLVSRTPQAKATAGEVLV 64
Db        8 IQAFNLNVGLSGISAIAERMNSLSKRVSLERHLGVLLRYSTRNVEPTGAGFFY 67

Qy        65 QAARKWVLQQ-----AETKA---QLSGRLAEI-PLTIANADSLSTFWPFVFNE-VASW 11
Db        68 KSAC--ASLDQNNAAPESVALRENDLCIGLRVMAPSFG-----TLWLGLVMFEFMARN 110

Qy        114 GGATLTRLEDAHTLSLLRRG-DVLGAATR-EANPVAGCEVVELGTMRHMLATALPSSLR 170
Db        120 PRLEVLIQDD--RIVDFEKGYDLAIRTRLPDSSLIAR----QLGTSRRVCCSPPEYL 172

Qy        171 DAYMVDGKL----DWAMPVLRFCKDVLDQRDLGDGRVDG-----PVGRRRVSIVPSA 216
Db        174 ERH---GPLRIEDILCHPCICYGSHNTPSQLWSFPVRVAGEPARMITPGRGFTNN---NG 222

Qy        220 EGFCGARIRGGWGILLPETQAAMLKAGEVI--LLDEIPIDITPMW 262
Db        228 QTMKDAAVRGLGLAMLFLFAEDAELAAGRIVEALPQERPLDLLII 272


RESULT 11
IQ10Y5           PRELIMINARY;                PRT;             292 AA.
AC               Q910Y5;
CD               Q910Y5;
DT      01-MAR-2001 (Tremblrel. 16, Created)
DT      01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE      01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE      PROBABLE TRANSCRIPTIONAL REGULATOR.
GN      PA2497.
OS      Pseudomonas aeruginosa.
OC      Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OX      Pseudomonas.
OY      NCBI_TaxID=287;
[1]     RN
RP      SEQUENCE FROM N.A.
RC      STRAIN=PA01.
RX      MEDLINE=20437337; PubMed=10984043;
RA      Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
```

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Golltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.,  
RT \*Complete genome sequence of *Pseudomonas aeruginosa* PA01, an  
RT opportunistic pathogen." ;  
RL Nature 406:959-964(2000).  
DR EMBL: AE004677; AAG05885.1; -;  
DR InterPro: IPR000847; -;  
DR Pfam: PF001066; -;  
DR PRINTS: PR00039; HTH\_1; 1.  
DR PROSITE: PS00044; HTH\_LYSR\_FAMILY; UNKNOWN\_1.  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
SQ SEQUENCE 292 AA; 32054 MW; 1B7E8AB9C4396CDF CRC64;

Query Match 10.4%; Score 152.5; DB 2; Length 292;  
Best Local Similarity 25.9%; Pred. No. 0.00049;  
Matches 84; Conservative 41; Mismatches 122; Indels 77; Gaps 14;

QY 4 IOLDTLLSIIDSGSPGASLALISPSAVSORVKALEHHVGRVLSRTQ-PAKATEAGEV 62  
DB 5 VLLRTFVSVDVDTGNFRAGEHLHLTQSTVSQQVIRLEQNLGCRLLDRSQRVLPTEGER 64  
QY 63 LVQAARKVLLQAETKQAOLSGRLAEIPLTAINADSLSTWFFPVEVASWGATLTLL 122  
DB 65 LLGYARRLLRLDEASEALSPAHGDLVGLVPEDLAGEVLPVLTFRTE---ERPLRL 121  
QY 123 EDEA----HTLSLLRGDVLGAVTREANPVAGCEVVELGTMRHLATATPSLRDAYMVDGK 178  
DB 122 EVESGLSHLLRLYRSGELDLLVQWGDSDC-----HARWAE----- 161  
QY 179 LDW---AAMPVLRFGPKVQDLDRDGLRVDGPGVR-RRVSIIVPSAEGFCE----- 224  
DB 162 LGWFGSAARPFGEESPEPVP-----LVVFPVGLYRQEMITHALESIGRRWRIGYSSAS 215  
QY 225 -----AIRRLGLWGLLP-----ETQAA-PMKAGCEVILLDEIPDTPMYWQW 266  
DB 216 LASLVAAVAGLGVSLPLGCVPEHRLGAGQAGPPPIAGLELALYARPELDS----- 268  
QY 267 RLRSRLA-RLTDVAVDAIEGLR 289  
DB 269 --AGRTLRLRLD-LCDARLEGILQ 289

RESULT 12  
Q9KM14 PRELIMINARY; PRT; 295 AA.  
AC Q9KM14;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE TRANSCRIPTIONAL REGULATOR, LYSR FAMILY.  
GN VCA0575.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
RT cholerae";  
RL Nature 406:477-483(2000).

DR EMBL: AE004388; AAF96477.1; -;  
DR TIGR: VCA0575; -;  
DR InterPro: IPR000847; -;  
DR Pfam: PF00126; HTH\_1; 1.  
DR PROSITE: PS00044; HTH\_LYSR\_FAMILY; UNKNOWN\_1.  
SQ SEQUENCE 295 AA; 32847 MW; 4BBD85F2A87B2F51 CRC64;

Query Match 10.4%; Score 152.5; DB 2; Length 295;  
Best Local Similarity 25.5%; Pred. No. 0.0005;  
Matches 79; Conservative 62; Mismatches 102; Indels 67; Gaps 17;

QY 1 MNPOLDTLLSIIDSGSPGASLALISPSAVSORVKALEHHVGRVLSRTQ-PAK 55  
DB 2 LSPITLEA-LHILDAIERRGFAAANELNRPSSLSYQIQKLEQDLDMIFDRSGHRAN 60  
QY 56 ATEAGEVLVQAARKVLLQAETK----AQLSRLAEIPLTAINADSLSTWFFP---VFN 108  
DB 61 FTEAGKLLERGR--AILAATEKLVNDATLLANGWELDTITIALDG-----IVPAANLFP 112  
QY 109 EVASWGATLT-LRUEDE--AHTLSLLR--RGDVLGAVTREANPVAGCEVVELGTMRHLA 163  
DB 113 MVEALGNISKTFVRIODEILACGWEALATGRADLLICPRIEALP-QDYKAETIGTKMIW 171  
QY 164 TATPS-----LRDAY-----MVDGKLDWAAMPVLRFGPKVQDLDRDGLRVDG 206  
DB 172 VAAPHYVRRSGEFEEAREKYRAIATADTAREQPAMSV-----NILQ----- 215  
QY 207 PVGRRVSIIVPSAEGFGEAIRGLWGLLPETQAAAPMLKAGCEVILL---DEIPDTPMYW 263  
DB 216 ---ROPRTVSNLDAKCKALVAGLGIGTLPLOVQAPYIDKGLKAIHGSELDMDIVLAW 272  
QY 264 QRWRL-ESRS 272  
DB 273 RNNQGEAKS 282

RESULT 13  
Q9KNH9 PRELIMINARY; PRT; 302 AA.  
AC Q9KNH9;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE TRANSCRIPTIONAL REGULATOR, LYSR FAMILY.  
GN VC2760.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
RT cholerae";  
RL Nature 406:477-483(2000).  
CC -!- SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.  
DR EMBL: AE004342; AAF95899.1; -;  
DR TIGR: VC2760; -;  
DR InterPro: IPR000847; -;  
DR Pfam: PF00126; HTH\_1; 1.  
DR PRINTS: PR00039; HTH\_LYSR.  
KW DNA-binding; Transcription regulation.  
SQ SEQUENCE 302 AA; 34281 MW; DDEEB28511225CE9 CRC64;

QY	7	DTL-----LSIIDSGSPGASLAIISPSAVSQRWKALEHHVGRVLSRTQPA-KATE	58
		: : :     : : :     : : :     : : :     : : :     : : :	
Db	25	DTLSEQFVLYLDVLDGSGSAAARKHPLTPSAVARMDALERAVGSTLLVTRTHAVRATP	84
QY	59	AGEVIVQARKWV-----LIQAEYKAQLSG--RLAEIPLTITAINADSLTWFPPVFNVEVAS	112
		: : :     : : :     : : :     : : :     : : :     : : :	
Db	85	AGLAFADARRIVTVELRLARAEAVSLSTAPQGLIRIDAPVPFGRRHLA---PAVADFKA	141
QY	113	WGGATITRLDEAHYLSLLRRGDVLGAV--TREANPVAGCEVW--ELGTMHRHLATATPS	168
		: : :     : : :     : : :     : : :     : : :     : : :	
Db	142	NPGLDVOLRLIGSFIDL---QGEHLGEVDVLVLRACGLPDSRLVATSLAPMVRVVCASPE	197

**THE UNIVERSITY OF CHICAGO**

**THE UNIVERSITY OF CHICAGO**

Db 261 VRELDDDELHTRYGIVSRAGYSLPLAAMID 294

Search completed: September 13, 2001, 17:56:49  
Job time: 299 sec